

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**

# **DataNotes — Um sistema colaborativo para anotação de estruturas de diretórios**

**Mariana Freitas de Gouveia**



Mestrado Integrado em Engenharia Informática e Computação

Orientador: Maria Cristina de Carvalho Alves Ribeiro (Professor Auxiliar)

Co-orientador: João António Correia Lopes (Professor Auxiliar)

28 de Fevereiro de 2013



# **DataNotes — Um sistema colaborativo para anotação de estruturas de diretórios**

**Mariana Freitas de Gouveia**

Mestrado Integrado em Engenharia Informática e Computação

Aprovado em provas públicas pelo Júri:

Presidente: Doutor Rui Carlos Camacho de Sousa Ferreira da Silva

Arguente: Doutor José Luís Brinquete Borbinha

Vogal: Doutor João António Correia Lopes

---

28 de Fevereiro de 2013



# Resumo

A investigação científica está cada vez mais baseada na recolha e exploração de quantidades apreciáveis de dados, o que leva os investigadores a considerar o seu depósito em repositórios. O objetivo destes repositórios é o armazenamento e preservação dos conjuntos de dados submetidos. Um problema que surge habitualmente é a dificuldade de interpretação destes dados. Para o ultrapassar, é necessário não só preservar os dados, mas também descrevê-los. As descrições podem ser mais ou menos especializadas, ou seja, podem apenas basear-se em descritores genéricos, como o título, a data e o criador, ou incluir características próprias de um domínio.

Atualmente, os repositórios que possuem as descrições mais ricas supõem a colaboração de um técnico — o curador de dados — que é responsável pela descrição dos conjuntos de dados. Neste cenário, como o investigador tem pouco controlo sobre a descrição dos dados, o processo tende a ser demorado. No caso do depósito e descrição dos dados em universidades, existem alguns projetos em que o processo de depósito recorre à figura do curador.

Este trabalho tem por objetivos a conceção e desenvolvimento de um sistema de anotação colaborativa, DataNotes, destinado ao uso por parte dos investigadores da Universidade do Porto. Os investigadores terão a possibilidade de depositar e descrever os seus conjuntos de dados autonomamente, recorrendo a algumas ferramentas que os auxiliam nesse processo. Poderão ainda descrever, em texto livre, outros factos observados que, através dos descritores disponíveis, não se conseguem anotar. Além dos investigadores terem a possibilidade de descrever os seus dados, podem também fazê-lo nos dados de outro investigador, desde que para tal estejam autorizados.

Com base na plataforma *Semantic MediaWiki* e uma das suas extensões, o *Semantic Forms*, foi desenvolvida uma nova extensão capaz de assegurar as funcionalidades do DataNotes. A plataforma foi também integrada com outro projeto denominado UPBox, para que possam ambos fazer parte de um processo de curadoria de dados científicos.

Os objetivos propostos foram cumpridos, embora haja algumas perspetivas de evolução futura. Possíveis melhorias incluem o desenvolvimento de uma nova extensão sobre a *Semantic MediaWiki* que permita que os esquemas de metadados sejam importados de forma automática para o DataNotes, assim como obter uma anotação semiautomática dos dados, através dos conteúdos dos conjuntos de dados a ser anotados.

**Palavras-Chave:** Gestão de dados de investigação, Anotação de dados, Esquemas de metadados, Sistema colaborativo, Wikis Semânticas.



# Abstract

Scientific research is increasingly based on the collection and use of significant amounts of data, which has led researchers to consider depositing them in data repositories. The goal of these repositories is the storage and preservation of datasets. One of the problems that usually arise is the difficulty experienced in interpreting these data. To overcome this issue, it is necessary not only to save the data but also to describe it. Descriptions can be specialised to varying degrees: they can be based only on generic descriptors such as title, date and creator, or include domain features.

The creation of rich descriptions for the datasets in a repository requires the collaboration of an information-management specialist who is responsible for creating those descriptions. In this scenario, as the researcher has little control over data descriptions, the process tends to be time-consuming. In the case of the storage and description of data in universities, there are some projects where the uploading process depends on these specialized staff, called “curators”.

This project aims to design and develop a collaborative annotation system to be used by researchers at the University of Porto. Using this system, they will be able to upload and describe their datasets themselves, using a set of tools to assist them in this process. They will also be able to describe, in free text, other observed facts that may be hard to fit into the available descriptors. Researchers are able not only to describe and update their own data but also data belonging to other researchers, provided they are authorised to do so.

Based on the Semantic MediaWiki platform and one of its extensions, Semantic Forms, this new extension was developed to ensure the functionalities of DataNotes. The platform was also integrated with another project called UPBox so that they can both be part of a complete research data curation process.

The initial objectives have been met, but there is plenty of opportunity for future expansion. Possible improvements include the development of a new extension on Semantic MediaWiki which would allow metadata schemas to be automatically imported into DataNotes, as well as semi-automatic annotation of data using the content of the datasets to be annotated.

**Keywords:** Research data management, Data annotation, Metadata schemas, Collaborative system, Semantic wikis.





# Agradecimentos

Gostaria de agradecer à Professora Cristina Ribeiro e ao Professor João Correia Lopes pela orientação prestada ao longo de toda a dissertação, assim como a disponibilidade demonstrada para o esclarecimento de dúvidas que foram surgindo.

Não poderia deixar de agradecer também ao estudante do programa doutoral em Engenharia Informática, João Rocha da Silva, uma vez que foi dele que surgiu parte da ideia da realização deste projeto e também foi demonstrando ao longo deste tempo grande disponibilidade que proporcionou um bom rumo neste projeto.

Um agradecimento também ao Professor Raul Vidal e ao Professor Augusto de Sousa que apostam numa formação de qualidade a todos os estudantes do MIEIC.

Para terminar, um especial agradecimento aos meus pais e irmãos que sempre me apoiaram na realização deste mestrado.

Mariana Gouveia



*“A ciência nunca resolve um problema sem criar pelo menos outros dez.”*

George Bernard Shaw



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contexto e enquadramento . . . . .	1
1.2	Motivação e objetivos . . . . .	2
1.3	Estrutura . . . . .	3
<b>2</b>	<b>Gestão de dados científicos</b>	<b>5</b>
2.1	Preservação dos dados . . . . .	5
2.2	Curadoria de dados . . . . .	6
2.3	Anotações de dados . . . . .	8
2.3.1	Metadados para dados de investigação . . . . .	8
2.3.2	Formatos de representação de metadados . . . . .	9
2.4	Plano de gestão de dados . . . . .	10
2.5	Análise de plataformas para repositórios . . . . .	11
2.5.1	DSpace . . . . .	12
2.5.2	Fedora . . . . .	13
2.5.3	Duracloud . . . . .	14
2.5.4	EPrints . . . . .	16
2.5.5	DataStaR . . . . .	16
2.6	Resumo e conclusões . . . . .	17
<b>3</b>	<b>Análise de plataformas wiki</b>	<b>19</b>
3.1	Os utilizadores e as wikis . . . . .	19
3.2	Funcionalidades padrão de uma wiki . . . . .	20
3.2.1	Gestão de autorizações . . . . .	20
3.2.2	Versionamento . . . . .	20
3.2.3	As extensões em wikis . . . . .	20
3.3	Wikis semânticas . . . . .	21
3.3.1	Representação de dados . . . . .	21
3.3.2	Metadados, Folksonomias e Ontologias . . . . .	22
3.4	Comparação de wikis semânticas . . . . .	23
3.4.1	MediaWiki e Semantic MediaWiki . . . . .	26
3.5	Resumo e conclusões . . . . .	29
<b>4</b>	<b>Anotação de conjuntos de dados apoiada numa wiki semântica</b>	<b>31</b>
4.1	Especificação de requisitos . . . . .	31
4.1.1	Atores . . . . .	32
4.1.2	Casos de uso . . . . .	32
4.2	Modelo concetual do domínio . . . . .	35

## CONTEÚDO

4.3	Resumo e conclusões . . . . .	36
<b>5</b>	<b>Arquitetura da extensão DataNotes</b>	<b>39</b>
5.1	Visão Geral . . . . .	39
5.2	Componentes . . . . .	40
5.3	Tecnologias . . . . .	42
5.4	Metodologias . . . . .	43
5.5	Resumo e conclusões . . . . .	43
<b>6</b>	<b>Implementação do DataNotes</b>	<b>45</b>
6.1	Base de dados da MediaWiki e Semantic MediaWiki . . . . .	45
6.2	Ficheiros base da MediaWiki e Semantic MediaWiki . . . . .	47
6.3	Criação de extensões na MediaWiki e Semantic MediaWiki . . . . .	47
6.3.1	Páginas especiais . . . . .	48
6.3.2	Hooks . . . . .	48
6.3.3	Ficheiro de configuração . . . . .	49
6.4	Extensão DataNotes . . . . .	49
6.4.1	Introdução de esquemas de metadados na wiki . . . . .	49
6.4.2	Criação de formulários com base em esquemas de metadados . . . . .	50
6.4.3	Criação de uma árvore com estruturas de ficheiros e diretórios da UPBox . . . . .	53
6.5	Resumo e conclusões . . . . .	57
<b>7</b>	<b>Demonstração do DataNotes</b>	<b>59</b>
7.1	Introdução . . . . .	59
7.2	Gestão dos esquemas de metadados . . . . .	59
7.3	Gestão das anotações . . . . .	64
7.3.1	Anotação de um ficheiro ou pasta . . . . .	64
7.3.2	Visualizar uma anotação . . . . .	64
7.3.3	Edição de uma anotação . . . . .	66
7.4	Testes . . . . .	67
7.5	Resultados e desempenho . . . . .	67
<b>8</b>	<b>Conclusões e Desenvolvimentos futuros</b>	<b>69</b>
8.1	Síntese . . . . .	69
8.2	Perspetivas e sugestões futuras . . . . .	70
	<b>Referências</b>	<b>73</b>
<b>A</b>	<b>Requisitos da extensão DataNotes</b>	<b>77</b>
<b>B</b>	<b>Interação UPBox e DataNotes</b>	<b>83</b>
B.1	API UPBox . . . . .	83
B.2	Projetos e estrutura de ficheiros de um utilizador . . . . .	84
B.3	Ficheiro RDF DataNotes . . . . .	86
B.4	Diagrama de interação UPBox e DataNotes . . . . .	87
<b>C</b>	<b>Guia de Instalação</b>	<b>89</b>

# Lista de Figuras

2.1	Base de utilizadores mundial instalada — DSpace ( <i>in</i> [Dur12a]) . . . . .	12
2.2	Conjunto de dados sem a extensão elaborada sobre o DSpace ( <i>in</i> [Uni12a]) . . .	13
2.3	Conjunto de dados com a extensão elaborada sobre o DSpace ( <i>in</i> [Uni12a]) . . .	14
2.4	Base de utilizadores mundial instalada — Fedora ( <i>in</i> [Dur12a]) . . . . .	15
2.5	Entrada de metadados — EPrints ( <i>in</i> [oS07a]) . . . . .	16
2.6	Arquitetura DataStaR ( <i>in</i> [SL10]) . . . . .	17
3.1	Páginas e links numa wiki na imagem superior, que apresentam conceitos e dados ligados por relações nas wiki semânticas na imagem inferior ( <i>in</i> [VKV <sup>+</sup> 06]) . .	22
3.2	Comparação de diversas características com a variedade de wikis semânticas existentes (Modificada: <i>in</i> [MJLP10]) . . . . .	25
3.3	Extensões da Semantic MediaWiki . . . . .	27
3.4	<i>Semantic Forms</i> — Criar propriedades, <i>templates</i> e categorias . . . . .	28
3.5	Projetos que usam a extensão SemanticForms . . . . .	29
4.1	Casos de utilização do DataNotes . . . . .	33
4.2	Modelo concetual do domínio antes da implementação da extensão DataNotes . .	35
4.3	Modelo concetual do domínio depois da implementação da extensão DataNotes .	36
5.1	Arquitetura SMW ( <i>in</i> [KVV <sup>+</sup> 07]) com a extensão DataNotes . . . . .	40
5.2	Diagrama de Componentes do DataNotes . . . . .	41
5.3	Pilha de tecnologias . . . . .	43
5.4	Metodologia de implementação . . . . .	44
6.1	Núcleo da base de dados da MediaWiki . . . . .	46
6.2	Categorias na base de dados da MediaWiki . . . . .	46
6.3	Esquema de metadados Dublin Core . . . . .	51
6.4	Classes das páginas especiais da extensão DataNotes . . . . .	52
6.5	Páginas especiais na wiki com a extensão DataNotes . . . . .	52
6.6	Página de visualização — O antes e o depois . . . . .	54
6.7	Classes da estrutura de diretórios e ficheiros da UPBox no DataNotes . . . . .	56
6.8	Estrutura de diretórios e ficheiros da UPBox no DataNotes . . . . .	56
7.1	Página principal — Wiki DataNotes da Universidade do Porto . . . . .	60
7.2	Login — Wiki DataNotes . . . . .	60
7.3	Criação de uma classe . . . . .	61
7.4	Criação de uma propriedade . . . . .	62
7.5	Criação de um <i>template</i> . . . . .	63
7.6	Criação de um formulário . . . . .	63

## LISTA DE FIGURAS

7.7	Criação de uma categoria . . . . .	64
7.8	Criação de uma anotação . . . . .	65
7.9	Introdução de novos esquemas de metadados para a anotação . . . . .	65
7.10	Introdução de novos descritores para a anotação . . . . .	66
7.11	Visualização de uma anotação com descritores de um esquema de metadados . . . . .	66
7.12	Visualização de uma anotação com descritores de dois esquemas de metadados . . . . .	67
7.13	Editar uma anotação . . . . .	68
7.14	Adicionar novos descritores a uma anotação em modo de edição . . . . .	68
B.1	Resposta ao pedido sobre o URL que devolve os projetos de um utilizador . . . . .	84
B.2	Código do pedido sobre o URL que devolve os projetos de um utilizador . . . . .	84
B.3	Resposta ao pedido sobre o URL que devolve a estrutura de um projeto . . . . .	85
B.4	Ficheiro RDF do DataNotes — Instância de metadados . . . . .	86
B.5	Diagrama de interação UPBox e DataNotes ( <i>in</i> [Bar13]) . . . . .	87



# Lista de Tabelas

A.1	UC1 — Login na wiki DataNotes . . . . .	77
A.2	UC2 — Selecionar ficheiro/pasta na UPBox . . . . .	78
A.3	UC3 — Inserir uma anotação . . . . .	78
A.4	UC4 — Visualizar uma anotação . . . . .	79
A.5	UC5 — Editar uma anotação . . . . .	79
A.6	UC6 — Eliminar uma anotação . . . . .	80
A.7	UC7 — Gerir grupos . . . . .	80
A.8	UC8 — Inserir um esquema de metadados . . . . .	80
A.9	UC9 — Inserir propriedade . . . . .	81
A.10	UC10 — Inserir <i>template</i> . . . . .	81
A.11	UC11 — Criar formulário . . . . .	82
A.12	UC12 — Criar categoria . . . . .	82
B.1	API UPBox — Receber ficheiro de anotação do DataNotes . . . . .	83
B.2	API UPBox — Obter projetos de um utilizador . . . . .	83
B.3	API UPBox — Envio de um ficheiro com as anotações em RDF . . . . .	84

## LISTA DE TABELAS

# Abreviaturas e Símbolos

AJAX	Asynchronous Javascript and XML
ANDS	Australian National Data Service
ASCII	American Standard Code for Information Interchange
CAVA	Communication Audio-Visual Archive
DANS	Data Archiving and Networked Services
DataONE	Data Observation Network for Earth
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
HTTP	Hypertext Transfer Protocol
ICPSR	Interuniversity Consortium for Political and Social Research
IEEE LOM	Institute of Electrical and Electronics Engineers — Learning Object Metadata
IEMSR	IE Metadata Schema Registry
JISC	Joint Information Systems Committee
LDAP	Lightweight Directory Access Protocol
MW	MediaWiki
NCBI	National Center for Biotechnology Information
NSF	National Science Foundation
OAR	Open Access Repository
OASIS	Open Access Series of Imaging Studies
OWL	Web Ontology Language
PHP	Hypertext Preprocessor
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
SF	Semantic Forms
SIGARRA	Sistema de Informação para Gestão Agregada dos Recursos e dos Registos Académicos
SMW	Semantic MediaWiki
SPARQL	SPARQL Protocol and RDF Query Language
UML	Unified Modeling Language
URL	Uniform Resource Locator
W3C	World Wide Web Consortium
XMI	XML Metadata Interchange
XML Schema	Extensible Markup Language Schema



# Capítulo 1

## Introdução

A web marcou um crescimento muito acelerado na quantidade e diversidade dos conteúdos gerados pelos indivíduos; por exemplo, todos colecionamos fotografias pessoais ou filmes. Também na ciência se tem vindo a assistir a um aumento dos dados gerados por sensores e outra instrumentação; por exemplo, as observações exportadas pelos instrumentos médicos podem ser automaticamente transferidas para um local de armazenamento e as estações meteorológicas criam fluxos de dados contínuos. Apesar de existirem já diversas comunidades, essencialmente académicas, a estudar formas de suportar a evolução dos dados, ainda há muito a avançar na área da gestão de dados.

Fazendo referência ao caso particular dos dados produzidos na área de investigação e embora o seu valor seja reconhecido, surgem vários problemas como a resistência à partilha, a quantidade e a dificuldade de os interpretar fora do contexto original. A forma como se pode contribuir para tornar visíveis os dados de investigação e fazer com que os investigadores colaborem com esta iniciativa é um caso ainda em estudo. Os projetos de vertente colaborativa necessitam que se contribua continuamente e que se invista no enriquecimento das plataformas colaborativas. Com base nestes incentivos podem nascer termos que descrevem diversos domínios que, mais tarde, servem de base à produção de vocabulários para a descrição dos dados. O processo tem tendência a ficar mais apelativo e os investigadores começam a sentir que a exploração dos dados e a sua reutilização são importantes. Os investigadores percebem também que a partilha dos dados faz com que estes cheguem às comunidades que os procuram e possam deste modo ser citados mais vezes.

### 1.1 Contexto e enquadramento

Para tornar possível a partilha e reutilização de dados de investigação, é necessária a presença de uma estrutura robusta capaz de recolher e preservar os diversos artefactos resultantes das atividades de investigação. Para fazer face ao crescente volume de conteúdo digital produzido em contexto de investigação, é necessário assegurar a sua gestão e preservação a longo prazo. Caso não sejam adotadas estas práticas, os dados correm o risco de se perderem devido às mais diversas

razões, de entre as quais se destacam os nomes que são atribuídos que deixam de fazer sentido e o formato do ficheiro que fica descontinuado. Grande parte dos investigadores defendem a criação de infraestruturas e práticas que lhes permitam ter acesso rápido e fácil a conjuntos de dados para o seu trabalho. Contudo, são também os primeiros a dizer que no seu trabalho não procuram locais seguros, como repositórios de dados, para guardar os dados e não disponibilizam descrições que podem ser úteis para a sua interpretação.

Além da importância de garantir que os dados permanecem acessíveis a longo prazo, é igualmente importante que sejam também interpretáveis. Quando se tem um conjunto de dados em mão, se não houver nada que os descreva, também não serão facilmente interpretáveis por outros investigadores que queiram usá-los. Ao nível internacional tem vindo a crescer a preocupação em estudar e melhorar as descrições dos dados. Embora existam dados vindos de domínios distintos, como por exemplo da engenharia, medicina e geologia, a sua descrição e organização são essenciais, tanto na preservação como na partilha entre investigadores. Os próprios investigadores têm de ser capazes de descrever corretamente os seus dados através de um mecanismo que se designa por anotação. Com o enriquecimento de anotações baseadas num domínio, as próximas anotações serão tendencialmente mais simples de produzir — considera-se que este processo de enriquecimento se assemelha a um processo de aprendizagem contínua.

## 1.2 Motivação e objetivos

A realidade de que a investigação científica tem vindo a evoluir ao longo dos últimos anos está cada vez mais presente em diversas comunidades de estudo. Isto leva a que haja uma maior preocupação sobre como lidar com alguns problemas existentes; entre eles o depósito, a descrição e a gestão dos dados de forma intuitiva — estas tarefas integram num processo de curadoria.

Atualmente, num processo de preservação e curadoria dos dados é necessário que os investigadores se façam acompanhar de curadores para os ajudarem durante o processo, algo que se pode tornar desmotivante por ser um processo moroso. O reforço destes recursos é dispendioso, pois existem áreas distintas: medicina, engenharias, geologia, ciências sociais, entre outras, onde é difícil que exista uma taxonomia capaz de responder a todo o vocabulário que lhes está inerente. Existem, porém, algumas comunidades que se preocupam em criar e enriquecer vocabulários dessas áreas através de estudos de aprendizagem contínua.

Com o propósito de ajudar a Universidade do Porto a minimizar os recursos necessários à implementação de uma solução de curadoria eficaz, o objetivo deste trabalho é o desenvolvimento de um sistema de anotação colaborativa destinado ao uso por parte dos seus investigadores. O sistema tem como objetivo mostrar aos investigadores que é um benefício despender alguns minutos a depositar e anotar os seus conjuntos de dados, para que futuramente possam ser mais referenciados, consultados e, acima de tudo, que se possa tirar o maior proveito deles. Para o concretizar, o sistema deve ser capaz de, num ambiente colaborativo, permitir a criação de anotações. Aqui existirão duas grandes vertentes em termos de anotação: os descritores mais comuns são já sugeridos e os outros poderão ser selecionados pelos investigadores. Embora a inspiração para este

projeto tenha surgido na Universidade do Porto, a necessidade de melhorar o processo de anotação de dados é geral, podendo tornar-se esta uma abordagem que pode ajudar não só investigadores da Universidade do Porto, mas também outros.

### 1.3 Estrutura

Este documento está dividido em oito capítulos: Introdução; Gestão de dados científicos; Análise de plataformas wiki; Anotação de conjuntos de dados apoiada numa wiki semântica; Arquitetura da extensão DataNotes; Implementação do DataNotes; Demonstração do DataNotes; Conclusões e Desenvolvimentos futuros.

No primeiro capítulo é feita a contextualização e o enquadramento do problema, são apresentadas as motivações envolvidas no projeto e os objetivos que lhe estão inerentes.

No início do segundo capítulo é elaborada uma descrição do estado da arte dos temas que o projeto aborda. Inicialmente é discutida a crescente necessidade de preservação de dados e posteriormente é introduzido o tema da curadoria de dados. A necessidade de descrição dos dados leva à exploração do conceito de anotação e metadados.

No terceiro capítulo foi preparada uma análise das plataformas wiki. Este estudo focou-se na apresentação de todas as potencialidades de uma wiki e nas vantagens da sua utilização para a construção do DataNotes. Duas das plataformas wiki foram especialmente detalhadas, uma vez que foram as selecionadas para servir de suporte a este projeto.

No quarto capítulo apresentam-se os requisitos e funcionalidades do projeto desenvolvido para anotação de conjuntos de dados. São também apresentados dois modelos conceituais do domínio, de forma a mostrar o que se pode encontrar com as plataformas escolhidas e aquilo que se pretende reaproveitar ou implementar de novo.

No quinto capítulo é descrita a arquitetura do DataNotes. Além da visão geral é ilustrado um diagrama de componentes que clarifica a arquitetura. Também são abordadas as tecnologias e as metodologias adotadas na implementação.

No sexto capítulo são descritos os detalhes de implementação da extensão DataNotes. Inicialmente faz-se referências às plataformas wiki que vão servir de base a este projeto. De seguida, para cada funcionalidade do DataNotes são descritos os pormenores da sua implementação e por fim, faz-se referência à integração do DataNotes com a UPBox.

No sétimo capítulo é apresentada uma pequena demonstração da wiki após a implementação do DataNotes. As funcionalidades base são descritas e acompanhadas de imagens que ilustram a interação.

Por fim, no último capítulo, são descritas as conclusões, apresentadas as perspetivas de desenvolvimento futuro e as contribuições deste trabalho.

## Introdução



## Capítulo 2

# Gestão de dados científicos

Neste capítulo é descrito o estado da arte do tema deste trabalho. Inicialmente é abordado o domínio mais abrangente, ou seja, a preservação e a curadoria dos dados. De seguida é introduzido o termo de anotações de dados e respetivas comunidades que estudam os esquemas de metadados, e também é feita pequena referência aos planos de gestão de dados que são elaborados. Por fim é feita uma análise de plataformas para repositórios de dados.

### 2.1 Preservação dos dados

A web é um meio rápido de aceder à informação, bem como o mais confortável. Através de pesquisas pode encontrar-se o que se pretende, embora nem sempre o mais desejado. Aquilo que é devolvido pelas pesquisas caracteriza-se como sendo relatórios, artigos, revistas, páginas web, entre outros.

Os dados são caracterizados como “... factos, números, letras e símbolos que descrevem um objeto, ideia, condição, situação ou outros fatores” [Bor11] e não são facilmente devolvidos pelas pesquisas. Uma das razões que leva a que isto aconteça é o facto de não existir o depósito dos dados num lugar da web. Isto levou à procura de novas alternativas que colmassem estes problemas relativos à preservação dos conjuntos de dados.

Borgman [Bor11] apresenta a motivação da partilha de dados de investigação. As políticas para promover a partilha de dados podem ser motivadas pelas necessidades de investigação ou pelo público em geral. Os investigadores podem ou não sentir-se motivados em partilhar os seus dados, quer seja dentro da sua comunidade ou para um público maior. Isto tende a acontecer porque os investigadores não sabem para que fins os dados são recolhidos e que tipo de abordagens podem ser feitas sobre eles. Apesar da partilha de dados ser difícil, isto não significa que ao longo do tempo esta ideia se mantenha. Os desafios são entender quais os dados que podem ser partilhados, por quem, com quem, em que condições e para que efeitos.

Supondo que os dados são partilhados, para chegar à preservação são necessários locais onde se possa armazená-los. O desenvolvimento das tecnologias possibilitou o estudo e o aparecimento de novos repositórios de dados. Existem comunidades, essencialmente relacionadas com

instituições de ensino, que não só colocam os dados nos seus repositórios, como também têm a preocupação de os preservar. No entanto, tal como refere no relatório técnico de estado da arte sobre repositórios científicos [RRFS10], existe um grande leque de repositórios que podem até ser encontrados em locais web, como o Databib<sup>1</sup> e o *Open Access Repository* (OAR) [Sim12] que os lista por áreas científicas distintas e multidisciplinares.

A preservação nestes repositórios é um processo contínuo onde os dados necessitam de ser geridos [CER11, DU11]. As medidas de preservação dos dados vão sendo modificadas e melhoradas ao longo do tempo; por exemplo, a tentativa de uniformizar os conjuntos de dados, ou seja, de colocá-los num só formato [RRL11].

## 2.2 Curadoria de dados

A curadoria de dados compreende o armazenamento, anotação e migração dos dados, entre outros processos, com vista à manutenção do acesso aos dados a longo prazo [DAM13]. O papel do curador e do investigador é fundamental para assegurar a consistência, a validade e a anotação sobre conjuntos de dados. Um investigador que pretende depositar um conjunto de dados com o resultado de uma experiência num repositório, dirige-se habitualmente a um curador que tem conhecimentos acerca de como pode anotar corretamente dados. Como resultado final, o curador poderá ainda ter que converter os dados para um formato aceite pelo repositório. Este trabalho de curadoria permite reduzir a duplicação e as inconsistências dos dados permitindo, a longo prazo, pesquisas de qualidade.

É possível que todo este processo seja feito conjuntamente ao longo de diversos passos; segundo um dos mais importantes centros de experiência em curadoria de informação digital, “... a curadoria e a preservação de dados são processos em curso, que requerem muita reflexão, investimento de tempo e recursos adequados” [DCC12]. O ciclo de vida da curadoria compreende diversos passos, que vão desde a conceção, criação, preservação, reavaliação, armazenamento, acesso e uso até à transformação, onde se criam novos objetos a partir do original.

De acordo com outra comunidade que é responsável pela curadoria de dados de investigação social e económica, o processo de curadoria está dividido em nove passos distintos: transferência de dados; atribuição de um tipo de processamento padrão; processamento dos dados; processamento da documentação; criação dos metadados; adição de informação relevante; publicação dos dados; entrega dos dados e preservação dos dados [Uni12b].

Um estudo elaborado pela Universidade do Porto e pela Universidade do Minho [RRFS10] revelou uma preocupação com a fraca robustez nas infraestruturas das organizações que lidam com grandes quantidades de dados científicos e por isso apresentou cinco formas distintas de abordagem à curadoria de dados:

- **Curadoria pelos cientistas ou técnicos que usam os dados** — Este tipo de processo encontra-se habitualmente em universidades ou centros de investigação onde não existe

---

<sup>1</sup><http://databib.org/>

nenhuma política institucional para a curadoria. São exemplos disto repositórios como *Communication Audio-Visual Archive*<sup>2</sup> (CAVA) e *Open Access Series of Imaging Studies*<sup>3</sup> (OASIS);

- **Curadoria por organizações científicas setoriais, instituições geradoras, recoletoras ou distribuidoras de dados** — Iniciativas como esta resultam em voluntariado por parte de universidades e instituições que habitualmente são financiadas por organismos que comecem a reconhecer a sua importância. A descrição dos dados nestes repositórios é mais detalhada e também para áreas específicas. São exemplos disto repositórios como *Data Archiving and Networked Services*<sup>4</sup> (DANS) e *National Center for Biotechnology Information*<sup>5</sup> (NCBI);
- **Curadoria por universidades ou centros de investigação** — Apesar deste ser um processo idêntico ao anterior, a iniciativa parte de uma universidade ou centro de investigação e, normalmente, contempla áreas distintas. Isto leva a que os curadores e os investigadores comuniquem frequentemente para conduzir este processo com sucesso. São exemplos disto repositórios como *DataShare*<sup>6</sup> e *EDINA*<sup>7</sup>;
- **Curadoria por organismos oficiais** — Tal como o nome indica, os organismos oficiais da gestão da ciência num determinado país fazem com que este tipo de iniciativa seja tomada. No entanto, o facto dos investigadores não estarem por perto e de existirem dados de diversas áreas nestes repositórios, leva a que a descrição não seja rigorosa. São exemplos disto repositórios como *Australian National Data Service*<sup>8</sup> (ANDS) e *Data Observation Network for Earth*<sup>9</sup> (DataONE);
- **Curadoria por comunidades informais em linha** — Este é um cenário que se tornará vulgar nos próximos anos com o aparecimento de comunidades que reúnem especialistas e amadores na recolha e descrição de conjuntos de dados de determinadas áreas. Existem sempre pessoas interessadas e especializadas capazes de fazer evoluir diversos tipos de taxonomias. É exemplo disto o repositório *Wikispecies*<sup>10</sup>.

Observa-se, através destes itens, que o processo de curadoria pode ser dirigido de forma distinta, ainda que a finalidade seja a mesma: assegurar existência de infraestruturas que permitem registar, descrever, pesquisar e aceder aos dados científicos.

---

<sup>2</sup><http://www.ucl.ac.uk/lscava/>

<sup>3</sup><http://www.oasis-brains.org/>

<sup>4</sup><http://www.dans.knaw.nl/en>

<sup>5</sup><http://www.ncbi.nlm.nih.gov/>

<sup>6</sup><http://datashare.is.ed.ac.uk/>

<sup>7</sup><http://edina.ac.uk/>

<sup>8</sup><http://www.ands.org.au/>

<sup>9</sup><https://www.dataone.org/>

<sup>10</sup><https://species.wikimedia.org/>

## 2.3 Anotações de dados

A ação de anotar dados tem lugar na produção de informação relacionada com um recurso ou na produção de metadados. No primeiro caso, a anotação é vista como um conjunto de comentários ou explicações adicionais elaboradas sobre conteúdo. No segundo caso, a anotação induz a utilização de modelos e normas para descrever os conteúdos. Aqui existe uma maior preocupação na procura de modelos e normas válidas que incluam descritores relacionados com os conteúdos a anotar.

A partir do momento em que existem repositórios que integrem *workflows* para um processo de curadoria de dados, a ação de anotar os dados pode fundamentar uma destas duas vertentes.

### 2.3.1 Metadados para dados de investigação

Para se descreverem dados pode optar-se pela utilização de modelos de metadados predefinidos que disponibilizam descritores que auxiliam na ação de anotar. Um dos modelos de metadados mais conhecidos e utilizados é o *Dublin Core*. O *Dublin Core* fornece um conjunto de 15 descritores que podem definir-se como gerais para a descrição de qualquer conjunto de dados de investigação.

Existem alguns projetos que têm diferentes formas de abordar o conceito de anotação. Os seguintes tópicos descrevem a perspetiva de dois projetos:

- O Open Annotation Collaboration (OAC) [HSS11] refere que a anotação permite aos utilizadores adicionar informação a recursos existentes, neste caso, aos conteúdos multimédia. Alerta ainda para a importância da interoperabilidade dos dados que são anotados.
- O projeto *Annotea* [KKPS02] refere que os tipos de anotação podem ser definidos pelos utilizadores, uma vez que estes têm diferentes pontos de vista e necessidades. É um projeto colaborativo baseado na web semântica, onde as descrições obtidas através da anotação são partilhadas. Uma anotação refere comentários, explicações, ou outros tipos de observações externas que podem ser ligadas a qualquer documento da web.

As anotações semânticas surgem nalguns projetos, como o *Annotea* [KKPS02], num contexto voltado para a web semântica e têm como finalidade integrar diversos significados de palavras e conseguir associar um significado a conteúdos na web.

A análise sobre a noção de anotação foi estudada a diferentes níveis, destacando nomeadamente as anotações em documentos e wikis semânticas. A ação de anotação em wikis semânticas tem a virtude de ter lugar num ambiente colaborativo onde vários investigadores podem intervir, o que leva à troca de informação entre os envolvidos [OMS<sup>+</sup>06].

Estes projetos mostram como a anotação de dados pode valorizar o trabalho de qualquer investigador, tentando adaptar os meios existentes ou até criar novos meios para a realizar corretamente. Ao elaborar a anotação, o investigador está efetivamente a construir uma instância de um modelo de metadados. Por outras palavras, os metadados podem ser definidos como “dados sobre dados”,

mas o termo é normalmente entendido como dados estruturados sobre recursos digitais que podem ser usados para ajudar a apoiar as pesquisas e a compreensão dos dados [JoB12].

### 2.3.2 Formatos de representação de metadados

O armazenamento dos metadados num formato definido para a sua representação pode ser feito de formas distintas. Vários autores estudam os formatos e esquemas a adotar para o seu projeto, mas a decisão recai a maioria das vezes sobre três linguagens de representação: *Extensible Markup Language* (XML), *Resource Description Framework* (RDF) e *Web Ontology Language* (OWL).

- O *Extensible Markup Language* (XML) [W3C08] pode ser usado para apresentar um modelo de metadados de uma forma simples. Por exemplo, para o esquema de metadados *Dublin Core* pode ser criada a tag <dc:title> para um título.
- O *Resource Description Framework* (RDF) [W3C04b] é uma versão do XML para interoperabilidade de metadados que define uma estrutura para triplos “Sujeito — Predicado — Objeto” através de um predicado chamado “rdf:tipo” que é usado para dizer que os conteúdos são de certos tipos.
- O *Web Ontology Language* (OWL) [W3C04a] acrescenta semântica a um esquema de metadados. Faz também uso dos triplos e permite, para além disso, definir e instanciar ontologias na web, baseadas em classes e propriedades. Por exemplo, pode indicar-se que “Se A é casado com B”, então isso implica “B é casado com A”. Com o uso desta linguagem, existe uma maior facilidade de interpretação por parte das máquinas.

Pelo facto de existirem estas diferenças entre diversos tipos de esquemas de linguagem existem algumas comunidades que se concentram no estudo da melhor forma de representação dos dados [Hun07]. Não existe uma resposta concreta sobre qual o esquema de linguagem mais correto a usar e compete a cada comunidade ou projeto determiná-lo. De seguida, são apresentadas as três comunidades mais conhecidas e ativas no desenvolvimento de modelos de metadados para anotação.

- A comunidade *Dublin Core Metadata Initiative* (DCMI) estuda modelos de metadados, proporcionando o acesso aberto a comunidades mundiais de pessoas que trabalham com os metadados. Toda a interação interpessoal promove a cooperação e interoperabilidade entre as normas e modelos. Grande parte dos projetos desenvolvidos na área são focados em modelos de metadados definidos pelo *Dublin Core Metadata Initiative* [PNN<sup>+</sup>07, Boa12];
- O projeto *IE Metadata Schema Registry* (IEMSR) desenvolve modelos de metadados como um serviço-piloto partilhado com o *JISC Information Environment*. Está essencialmente relacionado com os domínios de educação e bibliotecas digitais. Baseia-se em conjuntos de modelos de metadados como o *Dublin Core* ou o *IEEE Learning Object Metadata* (IEEE LOM) [Hun07, JIS09];

Os estudos desenvolvidos por comunidades como as anteriormente apresentadas têm uma vertente importante no desenvolvimento de esquemas de metadados e normas de diversos domínios, o que leva a que possam cada vez mais procurar-se soluções para a descrição dos dados que estão depositados nos repositórios.

### 2.4 Plano de gestão de dados

Um plano de gestão de dados é um documento formal que define como lidar com os dados, tanto durante a investigação como após a conclusão do projeto. O objetivo deste plano é considerar os vários aspetos da gestão de dados, da geração de metadados e da preservação dos dados, de forma a garantir que estes são bem geridos no presente e no futuro.

A elaboração de um plano de gestão de dados antes da recolha dos dados, garante que estes estão num formato correto, bem organizados e anotados. Isto economiza tempo, a longo prazo, porque não há necessidade de reorganizar ou reformatar os dados e os seus detalhes. No futuro, todos serão capazes de entender e usar corretamente os dados anotados. Os dados que são preservados têm o potencial de levar a novas descobertas e evitar a duplicação de estudos científicos.

Existem algumas agências de financiamento que estão a começar a exigir planos de gestão de dados como parte da proposta e avaliação do processo. É o caso da *National Science Foundation* (NSF) <sup>11</sup>. A política NSF inclui a partilha de resultados, de dados primários, de amostras físicas e de coleções. Esta política também menciona que a NSF impõe, através de uma variedade de mecanismos, o apoio e os incentivos adequados para a limpeza de dados, a documentação, a divulgação e o armazenamento. A NSF sugere que o plano pode conter:

- Os tipos de dados, as amostras, as coleções físicas, o software e outros a serem produzidos no decorrer do projeto;
- As normas a serem utilizadas para dados e metadados e formatos de conteúdo;
- As políticas de acesso e partilha, ou seja, de privacidade, confidencialidade, segurança, propriedade intelectual ou outros direitos de requisitos;
- As políticas e as provisões para reutilização, redistribuição e produção de derivados;
- As responsabilidades de gestão de dados atribuídos, isto é, que pessoas vão realmente ser responsáveis por assegurar a gestão de dados.

Muitas vezes é difícil e demorado para um grupo de investigadores ou mesmo para um departamento seguir uma abordagem coordenada, especialmente quando não há exemplares específicos a serem seguidos. Deste modo, foram criadas algumas ferramentas que auxiliam neste processo.

- A **DMPTool** <sup>12</sup> é uma ferramenta baseada na Web para criar e editar um plano personalizado. Esta ferramenta permite: (1) criar planos de gestão de dados prontos a usar para agências de

---

<sup>11</sup><http://www.nsf.gov/>

<sup>12</sup><https://dmp.cdlib.org/>

financiamento específicas; (2) atender aos requisitos para os planos de gestão de dados; (3) obter passo a passo as instruções e orientações para os dados do plano de gestão. Esta ferramenta foi desenvolvida por diversas universidades <sup>13</sup> para ajudar os utilizadores a atender a essas novas exigências por parte das entidades financiadoras.

- A **DMP Online** <sup>14</sup> é também uma ferramenta baseada na Web para criar e editar um plano personalizado. Esta ferramenta possui três funções principais: (1) ajudar a criar e manter diferentes versões dos planos de gestão de dados (versão de uso na candidatura, versão a ser desenvolvida durante o projeto e versão final); (2) fornecer orientações úteis em questões de gestão de dados e como atender aos requisitos de pesquisa dos financiadores; (3) exportar planos úteis numa variedade de formatos.

Para além de ferramentas existem também *templates* já desenvolvidos e pensados para serem utilizados.

- A Universidade de *Melbourne* desenvolveu um modelo de plano de gestão de dados <sup>15</sup>;
- A Universidade Nacional Australiana descreveu a gestão de dados em geral e incluiu um esboço de um plano genérico de gestão de dados <sup>16</sup>;
- O *Interuniversity Consortium for Political and Social Research* (ICPSR) disponibiliza links para exemplos de gestão dos planos de dados numa ampla variedade de disciplinas científicas <sup>17</sup>.

O *Joint Information Systems Committee* (JISC) <sup>18</sup> é também uma agência que financiou alguns projetos para ajudar os grupos de investigação, projetos e departamentos a cumprir melhores práticas disciplinares. Estes projetos exploram os desafios específicos da aplicação de políticas e diretrizes para necessidades específicas de uma gama de áreas de investigação. Através disto são formulados modelos de planos de gestão de dados e implementados sistemas de apoio para que os planos sejam executados.

Existem também outras agências financiadoras destes projetos <sup>19</sup>, cada uma focada em áreas específicas.

## 2.5 Análise de plataformas para repositórios

Algumas das plataformas para acesso a repositórios permitem ao investigador ou ao curador a ação de depósito e anotação dos dados. No entanto, grande parte dos investigadores tendem a adiar o processo para um momento mais oportuno.

---

<sup>13</sup>[https://dmp.cdlib.org/about/dmp\\_about](https://dmp.cdlib.org/about/dmp_about)

<sup>14</sup><http://www.dcc.ac.uk/dmponline>

<sup>15</sup>[http://www.eresearch.unimelb.edu.au/activities/research\\_data\\_management\\_for\\_researchers](http://www.eresearch.unimelb.edu.au/activities/research_data_management_for_researchers)

<sup>16</sup><http://anulib.anu.edu.au/training/>

<sup>17</sup><http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html>

<sup>18</sup><http://www.jisc.ac.uk/>

<sup>19</sup>[http://www.grants.gov/aboutgrants/agencies\\_that\\_provide\\_grants.jsp](http://www.grants.gov/aboutgrants/agencies_that_provide_grants.jsp)

## DSpace Markets

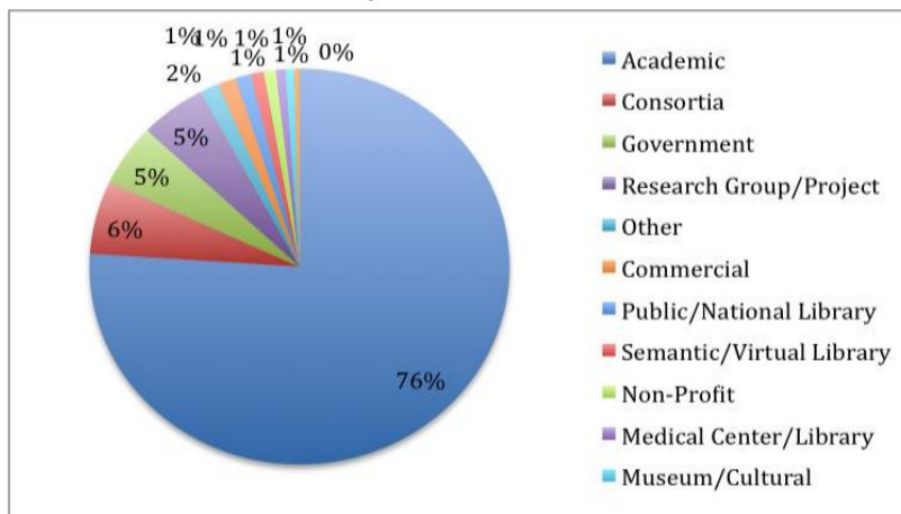


Figura 2.1: Base de utilizadores mundial instalada — DSpace (in [Dur12a])

O Duraspace [Dur12a] é uma fundação sem fins lucrativos que disponibiliza ferramentas de código aberto que ajudam as organizações a garantir que o seu conteúdo digital é durável e utilizável. É responsável por liderar 3 projetos (DSpace 2.5.1; Fedora 2.5.2; Duracloud 2.5.3) que se focam no desenvolvimento de tecnologias que promovem o acesso aberto e persistente dos dados digitais. Esta fundação colabora com comunidades académicas, científicas, culturais e tecnológicas de forma a garantir que, tanto atualmente como futuramente, estas comunidades tenham acesso a todas as suas coleções digitais. Além dos projetos já existentes, o Duraspace tem expandido novas áreas que incluem a exploração de estratégias sobre como lidar com a gestão dos dados e o desafio de converter a grande quantidade de dados produzidos em informação útil. Para além do Duraspace, outras entidades, nomeadamente académicas, possuem projetos que são bem sucedidos em determinados países, como é o caso do EPrints [Uni12c].

Nas sub-seções seguintes serão descritas as características gerais de algumas plataformas para acesso a repositórios disponíveis. Estas plataformas foram desenvolvidas para recolher, preservar e partilhar literatura científica, no entanto acabaram por anexar outros tipos de conteúdos digitais.

### 2.5.1 DSpace

O DSpace [Dur12b] é um software gratuito que preserva e permite o acesso fácil e aberto a todo o tipo de conteúdo digital, como por exemplo, texto, imagens, vídeos e conjuntos de dados. Existem, atualmente, cerca de 1400 instâncias da plataforma DSpace<sup>20</sup>, essencialmente em organizações académicas sem fins lucrativos e repositórios digitais abertos, como se pode verificar na Figura 2.1.

<sup>20</sup><http://www.dspace.org/whos-using-dspace/Repository-List.html>



488688.000000	38.76030000	-27.08409800	110.767	-53.750371
488689.000000	38.76029600	-27.08410600	110.813	-67.168032
488690.000000	38.76029500	-27.08410900	110.846	-80.584969
488691.000000	38.76029300	-27.08410800	110.871	-93.994527
488692.000000	38.76029200	-27.08410600	110.892	-107.391006
488693.000000	38.76029100	-27.08410500	110.911	-120.765010
488694.000000	38.76029000	-27.08410400	110.929	-134.106571
488695.000000	38.76029000	-27.08410300	110.944	-147.404566
488696.000000	38.76028900	-27.08410300	110.958	-160.644978
488697.000000	38.76028900	-27.08410300	110.969	-173.814814
488698.000000	38.76028900	-27.08410300	110.979	-186.901223
488699.000000	38.76028900	-27.08410300	110.988	-199.895619

Figura 2.2: Conjunto de dados sem a extensão elaborada sobre o DSpace (*in* [Uni12a])

Alguns exemplos de repositórios de dados que usam a plataforma DSpace são o *Edinburgh DataShare* da Universidade de Edimburgo, o MIT e a Universidade de Cambridge. Todos estes repositórios armazenam conjuntos de dados.

O projeto UPData [Uni12a, RRL11] teve como principal objetivo a exploração de uma solução para o repositório de dados da Universidade do Porto sobre o DSpace; este projeto é apenas um protótipo em exploração. Este protótipo determina as principais necessidades de curadoria de dados científicos em diversos domínios de investigação, de forma a melhorar a capacidade de curadoria, proporcionando a uniformização dos conjuntos de dados para a sua preservação e, com isto, conseguir uma pesquisa simplificada. O conjunto de dados referenciado na Figura 2.2 não passa de um conjunto de valores numéricos soltos e não interpretáveis. Contudo, a legibilidade da Figura 2.3 mostra como o projeto UPData conseguiu atribuir significado aos dados anteriormente perdidos.

Todo este processo adotado pelo UPData é interessante; no entanto, atualmente este depósito dos dados deveria ser acompanhado por um curador. Ainda assim, cada vez que existe um novo conjunto de dados o investigador passa conhecimento para o curador e este trata de construir e organizar esquemas capazes de descrever a informação desses conjuntos. Toda a informação base, como por exemplo, o título, a data, entre outros, têm já por base o uso do DC.

### 2.5.2 Fedora

Inicialmente desenvolvido pela Universidade de Cornell e atualmente como projeto Duraspace, o Fedora [SWP03] é uma arquitetura extensível que pode ser usada para desenvolvimento. É especialmente indicado para bibliotecas e arquivos digitais que contêm coleções complexas de materiais históricos, culturais e científicos, tanto para o acesso como para a preservação.

A comunidade Fedora [Com13] concentra-se na produção de ferramentas adicionais e aplicações que aumentam o repositório. Grande parte destas produções estão disponíveis para toda a comunidade em código aberto. Deste modo, o Fedora fornece:

- Um serviço de repositório central (com API bem definida);
- Um conjunto de serviços de suporte e aplicações;
- Suporte a RDF;

dc.creator	Bastos, Luísa; Deurloo, Richard
dc.title	Aerial Gravimetry Run (Processed Data)
dc.rights	open access
dc.type	Numerical Data
dc.date.issued	1992.0
dc.description	Final Results, combining the determined gravity values from the INS system with the GPS coordinates of the airplane in each moment

Sheet 0				
grav.gpstime	grav.latitude	grav.longitude	grav.altitude	grav.gravity.specific
488688.0	38.7603	-27.084098	110.767	-53.750371
488689.0	38.760296	-27.084106	110.813	-67.168032
488690.0	38.760295	-27.084109	110.846	-80.584969
488691.0	38.760293	-27.084108	110.871	-93.994527
488692.0	38.760292	-27.084106	110.892	-107.391006
488693.0	38.760291	-27.084105	110.911	-120.76501

Figura 2.3: Conjunto de dados com a extensão elaborada sobre o DSpace (in [Uni12a])

- Um software integrado com a tecnologia semântica *TripleStore*;
- Recursos que suportam a preservação digital.

Esta solução foi desenhada para apoiar uma variedade de necessidades do conteúdo digital, acabando por se tornar extremamente flexível. Esta flexibilidade permite-lhe também integrar com muitos tipos de sistemas corporativos e *web-based*. Desta forma, este repositório pode ser usado para suportar qualquer tipo de conteúdo digital e como consequência são vários os tipos de organizações que acabam por o utilizar. Esta solução oferece também a capacidade de consultar o repositório usando a linguagem de consulta SPARQL.

O Fedora tem uma base de utilizadores mundial instalada que inclui organizações académicas e de património cultural, universidades, instituições de pesquisa, bibliotecas universitárias, bibliotecas nacionais e órgãos governamentais, tal como pode ser visto na Figura 2.4.

O Duraspace concentra-se atualmente em melhorar o desempenho e estabilidade para todos os utilizadores do Fedora.

### 2.5.3 Duracloud

O DuraCloud [Dur13, Dur12a] é um serviço desenvolvido pelo DuraSpace que torna mais fácil para as organizações e utilizadores usarem os serviços na nuvem. Este projeto tem por base infraestruturas de nuvem já existentes para permitir maior durabilidade e acesso a conteúdo digital. Por outras palavras, o DuraCloud simplifica as transferências de cópias de conteúdo para a nuvem e posteriormente permite que sejam armazenadas em diferentes fornecedores de nuvem. Isto leva a que se torne mais simples mover e copiar conteúdos de qualquer formato e tamanho, ao contrário do que podia acontecer se apenas estivesse armazenado num só local.

Para dar suporte a esta estrutura estão continuamente a ser adicionados serviços que tornam o conteúdo do DuraCloud mais seguro e mais fácil de partilhar. Assim, o DuraCloud fornece essencialmente:

## Fedora Markets

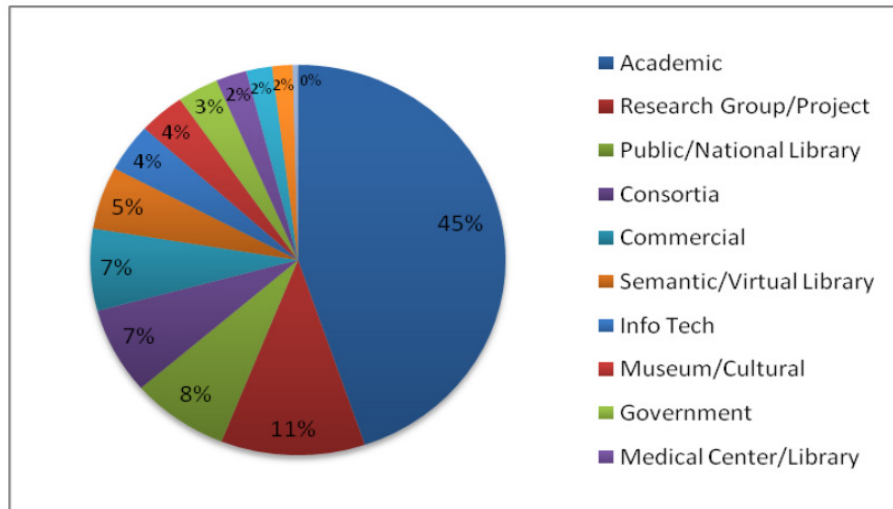


Figura 2.4: Base de utilizadores mundial instalada — Fedora (*in* [Dur12a])

- Serviços que permitem a preservação digital;
- Acesso a dados;
- Transformação e partilha de dados;

Por outro lado, o DuraCloud permite:

- **Backup e sincronização:** A ideia deste recurso baseia-se em copiar o conteúdo para vários fornecedores de armazenamento em nuvem e garantir que todas as cópias estão sempre sincronizadas.
- **Verificação:** A verificação dos arquivos armazenados através dos fornecedores de nuvem permite saber se o conteúdo está limpo e correto.
- **Vídeo e Streaming de Áudio:** Transmissão de vídeo e áudio através da nuvem, de forma a não existir a necessidade de sobrecarregar os computadores.
- **Exibição de Imagem:** Transmissão de imagens através da nuvem, de forma a não existir a necessidade de sobrecarregar os computadores.
- **Software Open Source:** Comunidade de código aberto, para quem pretenda utilizar o Duracloud tenha apenas de proceder ao *download*.

O DuraCloud é adaptado para qualquer pessoa, instituições individuais ou para múltiplas organizações.

Demo Repository eprints

Home | About | Browse by Year | Browse by Subject

Logged in as Dr Leslie Carr | [Manage deposits](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Logout](#)

Edit item: Article #112

You are both a [depositor](#) and an [editor](#) of this item. This is a depositor screen.

Type → Upload → **Details** → Subjects → Deposit

< Previous Save for Later Next >

★ Title

Abstract

★ Creators

	Family Name	Given Name / Initials	Email
1.	<input type="text"/>	<input type="text"/>	<input type="text"/>
2.	<input type="text"/>	<input type="text"/>	<input type="text"/>
3.	<input type="text"/>	<input type="text"/>	<input type="text"/>
4.	<input type="text"/>	<input type="text"/>	<input type="text"/>

Figura 2.5: Entrada de metadados — EPrints (in [oS07a])

## 2.5.4 EPrints

O EPrints [Uni12c] é um software gratuito e de licença aberta que partilha muitas das características observadas em sistemas de gestão de documentos. Ao longo dos anos tem vindo a crescer em termos de usabilidade e funcionalidades, sendo que atualmente é ainda utilizado como um arquivo aberto para depósito e gestão de conteúdos. Esta plataforma é essencialmente usada em repositórios institucionais e revistas científicas, numa gama de domínios bastante ampla.

Relativamente à preservação e curadoria dos dados, este repositório pode usar qualquer esquema de metadados. Dependendo do tipo de depósito que se cria, poderá decidir-se quais os descritores a incluir, como é referido na Figura 2.5. Uma pequena demonstração [oS07b] mostra que, depois de selecionado o tipo de ficheiro a armazenar e de o carregar, alguns descritores são preenchidos automaticamente.

## 2.5.5 DataStaR

O *Data Staging Repository* (DataStaR) [Cor12] é uma plataforma que pertence à *Cornell University Library* e é financiada pela *National Science Foundation*. Esta plataforma disponibiliza um conjunto de serviços destinados a apoiar a partilha de dados de investigação e a sua publicação.

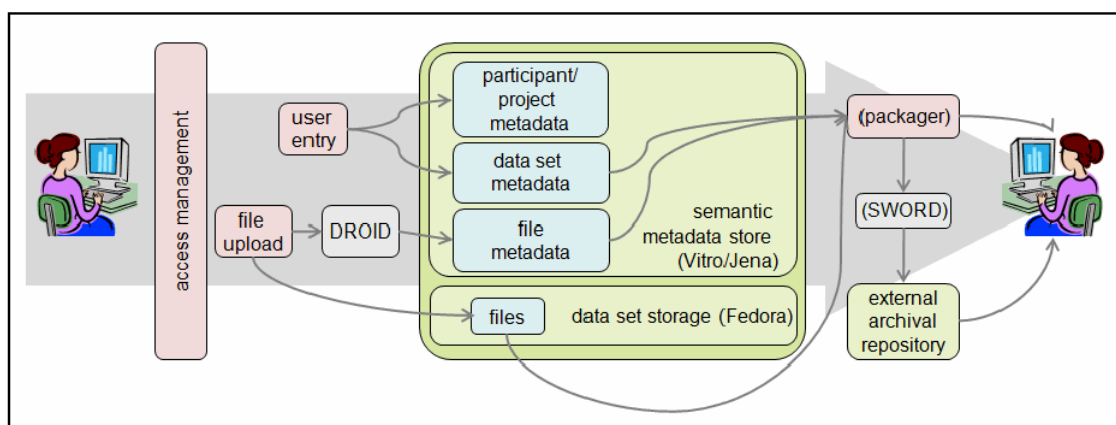


Figura 2.6: Arquitetura DataStaR (in [SL10])

Um investigador pode criar um conjunto de dados através do DataStaR com os metadados obrigatórios e carregá-lo para um repositório. Ou seja, os investigadores que desejam usar o DataStaR como uma plataforma para criar metadados podem posteriormente publicá-los num repositório externo suportado <sup>21</sup>. A seleção de um repositório para publicação desencadeia a apresentação de um esquema de metadados.

Apenas quatro elementos de metadados são exigidos como entrada ou seleção por parte dos investigadores: título, proprietário, permissões e repositório. Todos os outros elementos necessários ou são fornecidos automaticamente, ou têm um valor padrão. Ainda assim, o investigador pode adicionar outras descrições que pretenda. Esta plataforma adota uma abordagem web semântica de metadados no pressuposto de que as comunidades científicas cada vez mais a usam. Embora o DataStar ainda não suporte, a ligação dos dados será cada vez mais comum pois irá melhorar interoperabilidade e a descoberta através de recursos de informação [KCCr<sup>+</sup>10]. A forma como os metadados são tratados por esta plataforma leva a que possam ser reutilizados por um outro investigador.

A Figura 2.6 ilustra a arquitetura abordada por esta plataforma. O DataStaR consiste num repositório Fedora para o armazenamento de dados, um repositório de metadados semânticos com base em Viro/Jena, uma aplicação web semântica, e outros componentes de acesso livre: DROID, para identificação do formato do ficheiro e o SWORD para o seu depósito.

No entanto, todo este processo é auxiliado por curadores que pertencem à *Cornell University Library* e que gerem todos estes conteúdos.

## 2.6 Resumo e conclusões

Quando os dados científicos começaram a fazer parte do dia a dia dos investigadores, começou também a ser necessária a sua preservação em repositórios de dados. No entanto, um problema que surge habitualmente é a dificuldade na compreensão destes dados. A ação de anotação veio

<sup>21</sup> Repositórios: <http://datastar.mannlib.cornell.edu/index.jsp?primary=1900174249>

trazer a possibilidade de descrever os dados. A anotação pode ser mais ou menos rica, ou seja, pode apenas basear-se em descritores simples, como por exemplo, título, data e criador, ou podem ser mais focados no seu domínio. Através desta ação é gerada uma instância de um ou vários esquemas de metadados que englobam os conteúdos que foram anotados.

Nos diversos projetos analisados verificou-se que, para que este processo se torne mais completo, é ainda necessária a intervenção de um curador, dado que o funcionamento destes projetos exige um momento específico de depósito. No entanto, um aspeto positivo são as normas e os modelos de metadados que auxiliam neste processo.

Em suma, nenhum dos projetos analisados como plataformas existentes satisfaz os objetivos que se pretendem atingir com o DataNotes (a ação de anotação de forma autónoma) mas a preocupação e o estudo para evoluir nesse sentido está presente nas comunidades.

## Capítulo 3

# Análise de plataformas wiki

O DataNotes é um sistema colaborativo para anotação de estruturas de diretórios. O termo “colaborativo” induz uma característica que permite a um conjunto de pessoas participar, neste caso, num processo de descrição de dados. Existem diversas plataformas colaborativas e uma bem conhecida é a Wikipédia. As wikis são um *software* colaborativo e permitem a edição coletiva de documentos usando um sistema que não necessita que o conteúdo seja revisto antes da sua publicação.

Como potencial plataforma a usar para o desenvolvimento deste projeto, este capítulo apresenta uma análise das wikis e das suas funcionalidades padrão. Em maior detalhes são estudadas duas das plataformas, a *MediaWiki* e a *Semantic MediaWiki*, que apresentam características mais próximas do que se pretende utilizar para o desenvolvimento do DataNotes.

### 3.1 Os utilizadores e as wikis

O público alvo das aplicações produzidas compromete a escolha do tipo de sistema a desenvolver. Sendo o DataNotes destinado aos investigadores é necessário criar uma interface apelativa e fácil de usar por parte de engenheiros, arquitetos, psicólogos e médicos, entre outros.

Uma das principais características das plataformas wiki é a facilidade com que as páginas são criadas e alteradas. Através das wikis é possível tornar conhecimento público e geri-lo através da colaboração de pessoas. Em poucos anos, a wiki tornou-se uma ferramenta de interface simples e intuitiva, fácil de compreender e eficaz na gestão cooperativa de informação na web. A Wikipédia <sup>1</sup> é uma enciclopédia que usa uma plataforma wiki; a *Wikispecies* <sup>2</sup> é um catálogo de todas as espécies conhecidas até à data que também usa uma plataforma wiki; a *SNPedia* <sup>3</sup> recolhe dados de instrumentos médicos com informações genéticas dos humanos. As wikis são utilizadas para diversos fins e cabe a cada um decidir qual será a sua utilidade.

A constante dúvida por parte dos utilizadores das wikis é o facto dos conteúdos que pesquisam ou visualizam serem ou não verdadeiros. Talvez essa dúvida persista pelo modo de funcionamento

---

<sup>1</sup><http://www.wikipedia.org/>

<sup>2</sup><http://species.wikimedia.org/>

<sup>3</sup><http://www.snpedia.com/>

da Wikipédia, onde qualquer pessoa pode editar os conteúdos e facilmente se encontram páginas com conteúdo errado. No entanto, existem projetos wiki em comunidades onde parte dos conteúdos são validados por editores e outros podem ser editáveis pelo público.

## 3.2 Funcionalidades padrão de uma wiki

A utilização de wikis tem mais potencialidades do que aquilo que é visto pelos utilizadores; por exemplo, a gestão de utilizadores numa wiki é uma das tarefas que cabe a um administrador tratar e que não é visível ao resto dos utilizadores. Esta funcionalidade já está previamente implementada e pronta a ser utilizada. Isto acontece também com a gestão de versões.

As próximas subsecções descrevem como pode ser útil tirar proveito das potencialidades que são oferecidas pelas wikis.

### 3.2.1 Gestão de autorizações

Em muitos sistemas os administradores têm acesso a funcionalidades que os outros utilizadores não têm, como por exemplo, a gestão de utilizadores. Isto leva a que seja criada a aplicação cliente e um *backoffice* — estes desenvolvimentos levam tempo e têm custos associados.

As plataformas wiki já integram um sistema de gestão de utilizadores, onde é possível criar diversos grupos e atribuir-lhes diferentes permissões: esconder páginas, apenas dar permissões de leitura do conteúdo da página ou permitir ler e editar páginas, entre outros. Normalmente estas parametrizações são realizadas por um administrador.

### 3.2.2 Versionamento

O controlo de versões das páginas wiki é uma das principais características que se podem encontrar disponíveis neste tipo de plataforma.

Naturalmente, quando é efetuado qualquer tipo de alteração, a wiki guarda uma nova versão da página. Para cada página existe uma ligação que permite verificar o histórico das modificações, onde são listadas todas as alterações efetuadas. Aqui é possível fazer comparações entre duas versões selecionadas pelo utilizador e ainda repor uma versão antiga.

De um modo geral, o controlo de versões retira a todos os utilizadores o receio de modificar conteúdos. Esses conteúdos podem ser recuperados a qualquer instante.

### 3.2.3 As extensões em wikis

Existem algumas plataformas wiki que permitem incorporar extensões. As necessidades mais procuradas como acréscimos a uma plataforma wiki são expostas pelos utilizadores. São exemplos destas necessidades a importação de dados externos para páginas wiki, a construção de esquemas, a visualização de mapas, entre outros. Algumas comunidades ou indivíduos atendem a esses pedidos e acabam por implementar e partilhar o trabalho por eles desenvolvido. Além disso,



existem comunidades que se dedicam a implementar extensões para que a sua própria plataforma wiki seja uma das mais utilizadas, ainda que também possa receber contribuições externas.

### 3.3 Wikis semânticas

As plataformas wiki podem ser escritas em linguagens de programação como PHP, promovem a colaboração e a criação de páginas e destinam-se ao armazenamento de conteúdos produzidos pela ciência ou por outras áreas.

A web semântica tem como objetivo estender a web atual e estruturá-la para a tornar mais funcional. Ou seja, a web semântica relaciona significados de palavras e tem como finalidade conseguir atribuir um significado aos conteúdos através do uso de triplos (sujeito — propriedade — valor). Uma wiki semântica é um sistema que combina os dois termos anteriores, ou seja, a web semântica com as wikis [KMH08]. Permite a criação colaborativa, a edição e a interligação de páginas, mas também adiciona semântica sobre os dados introduzidos na wiki [SBBK08].

#### 3.3.1 Representação de dados

Uma wiki semântica está disponível numa linguagem formal, de modo a que as máquinas a possam processar. A notação formal pode ser introduzida nas páginas por utilizadores ou pode ser derivada a partir das páginas, nomes de página ou por meios de ligação.

As **categorias** que existem na Wikipédia servem para auxiliar a navegação entre páginas; mas, com as wikis semânticas foram introduzidos dois novos conceitos aliados às categorias: as **ligações** e os **atributos**.

- **Categorias** — Classificam páginas/artigos de acordo com seu conteúdo;
- **Ligações** — Classificam as ligações entre páginas/artigos de acordo o seu significado; podem também ser vistas como propriedades.
- **Atributos** — Especificam propriedades simples relacionados com o conteúdo de uma página/artigo.

A Figura 3.1 ilustra, na imagem superior, as páginas e respetivas ligações numa wiki como a Wikipédia. Na imagem inferior aparece a mesma página no entanto, os conceitos e os dados estão ligados por relações, tal como se pretende nas wikis semânticas. A página central está categorizada como “Capital” (ou seja, Londres), as ligações/propriedades estão representadas pelas setas e os atributos são os conteúdos para onde as setas apontam. Também se pode categorizar outra página nesta Figura como “País” (ou seja, Inglaterra) que tem uma propriedade “part of” que aponta para UK.

Com as características referidas, a plataforma é também capaz de implementar uma tecnologia, por vezes referida como **modelos**<sup>4</sup> semânticos. O mecanismo de modelo permite que os

---

<sup>4</sup>Do Inglês: Templates;

## Análise de plataformas wiki

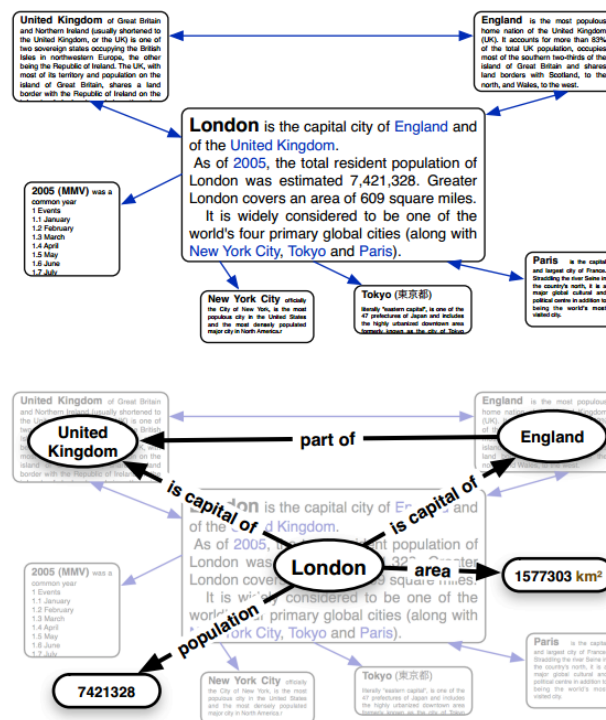


Figura 3.1: Páginas e links numa wiki na imagem superior, que apresentam conceitos e dados ligados por relações nas wiki semânticas na imagem inferior (in [VKV<sup>+</sup>06])

utilizadores incluam ligações e atributos predefinidos em novas páginas/artigos. Os modelos são páginas wiki cujo conteúdo é projetado para ser introduzido noutras páginas. Os modelos seguem uma convenção de que o nome é prefixado com “*Template:*”. Esta característica permite ter um formato fixo para variar o conteúdo e foi introduzido principalmente para assegurar uma distribuição consistente entre os artigos.

Por fim, podem ser atribuídas as categorias a estes modelos, pois tal como a definição diz, as categorias classificam páginas/artigos de acordo com seu conteúdo.

### 3.3.2 Metadados, Folksonomias e Ontologias

A web semântica é a base para a estruturação e interoperabilidade dos recursos na web. As wikis semânticas necessitam também de estruturar os seus dados para a definição dos modelos.

O *World Wide Web Consortium* <sup>5</sup> (W3C) propõe normas e padrões para a construção de vocabulários e definição de regras para gestão de dados. As ontologias suportam a evolução de vocabulários que podem definir relações entre diferentes conceitos. A ontologia desenvolvida é composta pelo modelo de metadados, por exemplo, *Dublin Core*, mais os metadados das *folksonomias*. A *folksonomia* [CB10] é o resultado da atribuição livre e pessoal de etiquetas <sup>6</sup>, pelos próprios utilizadores, às informações ou objetos. A *folksonomia* é também um termo que reúne

<sup>5</sup><http://www.w3.org/>

<sup>6</sup>Do Inglês: Tagging;

atualmente temáticas em exploração que pretendem que cada página web contenha metadados legíveis pelas máquinas que descrevem o seu conteúdo [Siq12]. Isto permite melhorar a precisão das pesquisas dos motores de pesquisa como o *Google*.

A dificuldade reside em saber como levar uma enorme variedade de autores de páginas web a adicionar metadados consistentes e confiáveis nas suas páginas. O conceito de modelo de metadados controlados, como *Dublin Core* e outros ainda mais complexos, contraria ainda esta ideia. A longo prazo é importante a contribuição com conteúdo semântico.

A definição dos modelos das wikis semânticas pode basear-se na definição ou utilização de vocabulários, quer sejam normas, modelos de metadados ou *folksonomias*. Com o DataNotes pretende-se contraditar um pouco os estudos rígidos de vocabulários interoperáveis e recorrer às *folksonomias*.

### 3.4 Comparação de wikis semânticas

As wiki semânticas têm características que as fazem distinguir de outras plataformas. Uma são mais conhecidas do que outras porém, a decisão sobre qual se deve utilizar depende totalmente das necessidades pretendidas. De seguida são apresentadas as quatro características que mais se destacam como perspectiva de possível utilização por parte do DataNotes:

- **Ambiente colaborativo** — Um ambiente colaborativo permite que um investigador possa anotar conjuntos de dados associados aos seus projetos e aos projetos onde está inserido;
- **Preservação** — Nos sistemas de anotações são descritos dados que estão armazenados, normalmente, em repositórios. Assim como o repositório é responsável por armazenar os dados, a plataforma wiki deve preservar as suas descrições para que possam ser consultadas e editadas;
- **Redução de dependências** — As wikis são instaladas, por quem as desenvolveu, numa máquina virtual que disponibiliza um URL para acesso através de um *browser*. Um utilizador que pretenda consultar a wiki terá de aceder a um *browser* e indicar o respetivo URL. O facto de não ser necessário instalações no seu computador pessoal torna a utilização da wiki mais ágil e com um nível de dependências muito reduzidas. O utilizador apenas necessita do acesso à Internet para conseguir navegar na wiki.
- **Utilização intuitiva para o utilizador** — Entre os investigadores e curadores, o leque de diversidade de conhecimentos informáticos é diverso, sendo que uns poderão dominar bem qualquer tipo de sistemas e outros poderão abandonar a utilização do sistema caso a curva de aprendizagem seja demasiado íngreme. As plataformas wiki necessitam de pequenas noções sobre como devem ser utilizadas, que são facilmente adquiridas por tutoriais disponível na web.

Um estudo realizado em 2010 [MJLP10] revelou dados importantes e esclareceu algumas dúvidas relativamente às características entre diversas wikis semânticas, como é possível verificar

na Figura 3.2 (os pontos preenchidos têm o significado de “suportado”, enquanto que os pontos não preenchidos têm o significado contrário). A tabela encontra-se dividida em três partes e cada uma destas partes está dividida em sub-partes.

Iniciando a análise desta tabela pela coluna “ferramentas de web semântica”, pode observar-se que as diferentes wikis semânticas têm algumas ferramentas em comum. A forma de representação da linguagem e o modo como os dados são guardados são as sub-partes de maior importância. As linguagens de representação mais comuns são *Ontology Web Language* (OWL) e *Resource Description Framework* (RDF). O armazenamento dos conteúdos é guardado numa base de dados em forma de texto das páginas wiki ou em formato RDF. As vantagens da utilização do segundo modo reside na organização dos conteúdos e posterior pesquisa mais eficiente. A exportação de uma página wiki para uma ontologia é um outro exemplo de uma ferramenta semântica, que permite a exportação dos conteúdos armazenados na wiki.

De forma a facilitar a interação com a wiki são diversas as funções que a integram. A gestão de utilizadores e de versões são imprescindíveis. A sintaxe utilizada deve ser de fácil interpretação para não desencorajar a sua utilização. O uso de **formulários** numa wiki atrai aqueles que não estão habituados a lidar com a sua sintaxe. Os formulários são um conjunto de caixas de texto nas páginas wiki onde podem ser introduzidos conteúdos. Cada caixa de texto destina-se a um conteúdo específico, por exemplo, à introdução de um nome.

Por fim, a informação sobre o estado da wiki permite saber se existem versões estáveis de cada plataforma. Para além disso, é necessário ainda apurar se a empresa ou instituição que desenvolveu o produto continua ou não disponível.

As características anteriormente detalhadas procuraram dar destaque às sub-partes que teriam uma maior importância para a implementação do DataNotes. Assim, caso se pretendesse seleccionar uma destas plataformas para servir de suporte ao desenvolvimento do DataNotes, apenas duas das doze wikis semânticas seriam escolhidas: *Knowledge in a wiki* (KiWi) e *Semantic MediaWiki Plus* (SMW+). Embora a *OntoWiki* seja também uma plataforma forte, esta perde por não ter uma sintaxe definida. Todas as outras apresentam pequenas falhas, ou porque não fazem uso de formulários ou porque não têm controlo de versões.

Após esta análise é possível concluir que existem plataformas que são mais completas do que outras. No entanto, da escolha final devem fazer parte as funcionalidades mais requeridas para o sistema que se pretende implementar. As diferenças observadas entre a KiWi e a SMW+ não permitem decidir qual poderia ser a plataforma base para o DataNotes.

Através da análise particular de ambas as plataformas, podemos concluir que, a KiWi <sup>7</sup> propõe uma abordagem para a gestão do conhecimento, ou seja, combina a filosofia da wiki com a inteligência e métodos da web semântica. A KiWi é uma plataforma para a implementação e integração de diferentes tipos de serviços de software social, como por exemplo *blogs* e redes sociais. Apesar de apresentar potencialidades, esta wiki não vai ao encontro do que se pretende para a implementação do DataNotes.

---

<sup>7</sup><http://www.kiwi-project.eu/>

## Análise de plataformas wiki

Semantic Web tools													Wiki functions					Status		
Repre- sentation language	Ontology export	Ontology import	Query system	SparQL access	Inference engine	Data storage	Annota- tion type	Syntax	WYSI- WYG editor	Forms	Other editing mode	Version manage- ment	Users manage- ment	Editor	License	Language	Version			
AceWiki	OWL, SWRL	○	○	Text, advanced queries	○	Pellet	Text	ACE	○	○	Predictive editor	○	○	University of Zurich	LGPL	Java	alpha			
BoWiki	OWL	●	OBO format	Text, advanced queries	○	Pellet	DB	wiktext	○	○	○	●	●	Max Planck Institute	GPL	PHP	stable			
KiWi	RDF, OWL	●	○	Text, SparQL, advanced queries	●	Jena	RDFstore + DB	wiktext	●	●	Auto- completion	●	●	KiWi European Project	BSD	Ajax, Java	stable			
Knoodl	Unknown	●	●	Text, SparQL, advanced queries	●	○	RDFstore	wiktext	●	○	○	●	●	Revelyix	Unknown	Java	stable			
KnowWE	OWL, d3web XML	●	d3web format	Text, SparQL	○	Sesame	○	wiktext	○	●	Auto- completion	●	●	University of Würzburg	LGPL	Java	beta			
OntoWiki	RDF(S), OWL	●	○	Text, SparQL, advanced queries	●	pOWL	RDFstore	Un- know	●	●	Auto- completion	○	●	University of Leipzig	GPL	PHP	stable			
Semantic MediaWiki	RDF(S), OWL	●	○	Text, SparQL, advanced queries	○	○	DB	wiktext	○	○	○	●	●	University of Karlsruhe	GPL	PHP	stable			
SMW+	RDF(S), OWL	●	○	Text, SparQL, advanced queries	●	KAON 2	DB, RDFstore	wiktext	●	●	○	●	●	Ontoprise GmbH	GPL	PHP	stable			
Subleme	RDF	○	○	Text	○	○	HTML pages	wiktext	●	○	○	○	●	Unknown	GPL	Unknown	Unknown			
SWim	RDF(S), OWL	●	OMDoc format	Text, SparQL, advanced queries	○	○	DB	wiktext	●	●	○	●	●	KWAPK research group	GPL	Java, XSLT	alpha			
TaOPis	F-Logic	●	○	Text, Flora-2, advanced queries	○	Flora-2	DB	wiktext	○	○	○	●	●	University of zagreb	GPL	Pytho n, PHP	alpha			
Wikidmart	Unknown	○	○	Text, SparQL, advanced queries	●	●	DB	wiktext	○	●	○	●	●	Zagile, inc	AGPL	PHP	stable			

Figura 3.2: Comparação de diversas características com a variedade de wikis semânticas existentes (Modificada: *in* [MJLP10])

A SMW+ representa a plataforma *Semantic MediaWiki* (SMW) com algumas das suas extensões. A Figura 3.2 mostra apenas as características na versão base das plataformas wiki. A utilização da SMW e posteriormente a adição das extensões que se pretende utilizar, é uma melhor abordagem a seguir. A subsecção 3.4.1 explora com mais detalhe a SMW e a plataforma sobre a qual acenta, a *MediaWiki*.

### 3.4.1 MediaWiki e Semantic MediaWiki

A *MediaWiki* [Bar08, Kor12] é uma plataforma wiki escrita em PHP e usa uma base de dados *MySQL*, *PostgreSQL*, *Microsoft SQLServer*, *IBM DB2* ou *Oracle* como suporte ao armazenamento dos conteúdos. Caracteriza-se como uma plataforma madura, estável e segura. Servindo de suporte à Wikipédia, a sua utilização é bastante intuitiva e tem um menu lateral facilmente personalizável. É também uma plataforma facilmente extensível e existem diversas contribuições para o seu crescimento. Além das funcionalidades comuns já referidas na secção 3.2, esta plataforma permite que um utilizador possa ser informado das alterações ao conteúdo das páginas que estejam sob a sua observação <sup>8</sup>.

Uma das wikis semânticas mais conhecidas e utilizadas é a SMW; atualmente existem mais de 300 ativas em todo o mundo. A SMW é uma extensão da MW tradicional que usa tecnologias da web semântica, para possibilitar que no ambiente wiki, além da navegação por meio de ligações, possa também ter uma estrutura interpretada por máquinas. Junto de quem a procura, esta plataforma fornece um grande número de extensões que satisfazem diversos requisitos.

A SMW proporciona a possibilidade de desenvolver uma extensão em separado para o Data-Notes e futuramente, poderá ser mais fácil estender e implementar novas extensões para o âmbito das anotações — foi portanto esta a plataforma selecionada para esta implementação.

A SMW inclui novos conceitos que até agora eram desconhecidos, como por exemplo propriedades e notações. A criação de propriedades semânticas surge através da introdução de pequenas notas no conteúdo de um artigo que referencia outro. Se uma página tem uma ligação para outra, a esta ligação pode ser atribuído um significado. Ou seja, se existir uma página que descreve um livro onde existe uma ligação para uma outra página que descreve um autor, pode ser criada a relação semântica “O livro foi escrito por um autor”. Cada propriedade tem um tipo associado, por exemplo, uma página, um conjunto de caracteres <sup>9</sup>, um número, um *Uniform Resource Identifier* (URI), entre outros. Para suportar estas propriedades, a base de dados da MW sofreu ligeiras transformações — existem tabelas que guardam as propriedades por tipo.

Embora exista já o suporte semântico nesta plataforma, ao longo do tempo surgem necessidades que originam desenvolvimentos de novas extensões que, mais tarde, integram na SMW. A Figura 3.3 <sup>10</sup> ilustra as inúmeras extensões que foram implementadas até à data. Estas extensões estão divididas em dez grupos distintos e cada grupo acrescenta valor a esta plataforma; as extensões são desenvolvidas por diversos autores e são categorizadas por estes grupos. No entanto,

---

<sup>8</sup>Do Inglês: watchlist

<sup>9</sup>Do Inglês: string

<sup>10</sup>Disponível em: [http://semantic-mediawiki.org/wiki/Help:SMW\\_extensions](http://semantic-mediawiki.org/wiki/Help:SMW_extensions)

## Análise de plataformas wiki

1 Adding and modifying data	
1.1 Semantic Forms	←
1.2 Semantic Image Input	
1.3 Halo	
1.4 Rule Knowledge	
1.5 Collaboration extension	
1.6 WYSIWYG	
1.7 MetaVidWiki	
1.8 Semantic Forms Inputs	
1.9 Semantic Forms Select	
1.10 Semantic Signup	
1.11 Semantic Social Profile	
1.12 Semantic Glossary	
2 Searching and browsing	
2.1 Enhanced Retrieval	
2.2 Frequent Pattern Tag Cloud	
2.3 Semantic Drilldown	
2.4 Semantic Gardening	
2.5 Semantic Web Browser	
2.6 SolrStore	
2.7 TreeView	
3 Displaying data	
3.1 Semantic Result Formats	
3.2 Semantic Maps	
3.3 Semantic Compound Queries	
3.4 Semantic graph extensions	
3.5 Semantic Project Management	
4 Storing additional data	
4.1 Semantic History	
4.2 Semantic Internal Objects	
4.3 Semantic Extra Special Properties	
4.4 JSON	
5 Storing data via an RDF triplestore	
5.1 LinkedWiki	
5.2 RDFIO	
5.3 SparqlExtension	
5.4 Triple Store Connector	
6 Protecting data	
6.1 Access Control List	
6.2 Semantic Access Control	
6.3 SemanticACL	
7 Awareness and workflow	
7.1 Semantic Notifications	
7.2 Semantic Tasks	
7.3 Semantic Watchlist	
8 Importing data	
8.1 Data Import	
8.2 Data Transfer	
8.3 External Data	
8.4 Rich Media	
9 Exporting data	
9.1 LightweightRDFa	
10 Utilities	
10.1 Semantic Tagbox	
10.2 SemanticUpdateOnPurge	
11 Obsolete or discontinued extensions	
11.1 FolkTagCloud	
11.2 Freebase	
11.3 Question API	
11.4 RDFa	
11.5 Semantic Calendar	
11.6 Semantic Gallery	
11.7 Semantic Google Maps	
11.8 Semantic Layers	
11.9 Semantic NotifyMe	
11.10 SemanticQueryFormTool	
11.11 SemanticQueryRDFS++	
11.12 SMW OntologyEditor	
11.13 Semantic Incoming Properties	

Figura 3.3: Extensões da Semantic MediaWiki

## Análise de plataformas wiki

The figure consists of three numbered screenshots of the Semantic Forms extension interface:

- Criar uma propriedade (1):** This screenshot shows the form for creating a new property. The 'Nome da propriedade' field contains 'Was published in year'. The 'Tipo' dropdown menu is open, showing options like 'Página', 'Cadeia', 'Texto', 'Código', 'Variável', 'Booleano', 'Número', 'Coordenadas geográficas', 'Temperatura', 'Data', 'Email', 'URL', 'Anotação-URI', 'Telephone number', 'Record', and 'Quantity'. The 'Número' option is selected. Below the dropdown, there are buttons for 'Gravar página' and 'Antevisão'.
- Criar uma predefinição (2):** This screenshot shows the form for creating a new template. The 'Nome da predefinição' field contains 'Book'. The 'Categoria definida pela predefinição (opcional):' field is empty. Below this, there is a section titled 'Campos da predefinição:' with instructions. Three example fields are shown: 'Authors', 'Year', and 'Number of pages'. Each field has a 'Mostrar etiqueta:' field, a 'Propriedade semântica:' dropdown, and a checkbox for 'Este campo permite uma lista de valores, separados por vírgulas'. There are 'Apagar' buttons for each field. At the bottom, there is an 'Adicionar campo' button.
- Criar uma categoria (3):** This screenshot shows the form for creating a new category. The 'Nome da categoria' field contains 'BookCategory'. The 'Formulário padrão:' dropdown menu is open, showing options like 'Book' and 'Geografias'. Below the dropdown, there are buttons for 'Gravar página' and 'Antevisão'.

Figura 3.4: *Semantic Forms* — Criar propriedades, *templates* e categorias

apenas a extensão *Semantic Forms*, que pertence ao grupo 1 de “Adicionar e modificar dados” assinalado na Figura 3.3, poderá ser utilizada para auxiliar funcionalidades do DataNotes. O *Semantic Forms* destina-se a ser usado para dados estruturados que têm marcação semântica e faz uso de formulários para adicionar, editar e consultar dados na wiki, tal como se pretendia para o desenvolvimento do DataNotes. Assim, esta extensão permite a criação e edição de **propriedades**, **templates** e **categorias** através de formulários.

A título ilustrativo, a Figura 3.4 mostra o *layout* para a criação de propriedades(1), *templates*(2) e categorias(3) com um exemplo prático: a criação de um formulário para um livro. Inicialmente são criadas todas as propriedades que se pretende adicionar, de seguida é a vez de criar



## Análise de plataformas wiki

<b>Culture:</b> <ul style="list-style-type: none"><li>• <a href="#">Oh Internet</a></li></ul> <b>Politics and government:</b> <ul style="list-style-type: none"><li>• <a href="#">Region 3 Leves wiki</a></li><li>• <a href="#">Politique.sémantique.com</a> (in French)</li><li>• <a href="#">Verwaltungskooperation</a></li><li>• <a href="#">UNDP Inter-Municipal Cooperation</a></li></ul> <b>Place listings:</b> <ul style="list-style-type: none"><li>• <a href="#">Hacker Spaces</a></li><li>• <a href="#">Mikomos</a></li><li>• <a href="#">World Directory of Churches</a></li></ul>	<b>Goods and services:</b> <ul style="list-style-type: none"><li>• <a href="#">WikiProducts</a></li><li>• <a href="#">Textbook Revolution</a></li><li>• <a href="#">Dexid</a></li><li>• <a href="#">Free Software Directory</a></li></ul> <b>Science and technology:</b> <ul style="list-style-type: none"><li>• <a href="#">Domotiki.eu</a> about Home Automation (in French)</li><li>• <a href="#">OpenEI: Open Energy Information</a></li><li>• <a href="#">SNPedia</a></li><li>• <a href="#">Technical Presentations</a></li><li>• <a href="#">Life encyclopedia</a> (in Lithuanian - Gyvosios gamtos enciklopedija)</li></ul>	<b>History:</b> <ul style="list-style-type: none"><li>• <a href="#">Familypedia</a></li></ul> <b>Individual organizations:</b> <ul style="list-style-type: none"><li>• <a href="#">Creative Commons Wiki</a></li><li>• <a href="#">OLPC Wiki</a></li></ul> <b>Environment:</b> <ul style="list-style-type: none"><li>• <a href="#">Milieuhulp</a></li></ul>
---	--	---

Figura 3.5: Projetos que usam a extensão SemanticForms

o *template* indicando o seu nome, atributos e respetivas propriedades. Após este passo, embora não esteja indicado na Figura 3.4 é necessário criar um **formulário** com o nome de um livro, por exemplo “Book” e atribuir-lhe o *template* acabado de criar. Um formulário resulta num modelo que contém descritores e propriedades predefinidas. Por fim, pode ainda atribuir-se ao formulário criado uma nova categoria. É importante que não se confunda os formulários (caixas de texto para introdução de conteúdos) com a criação de um formulário na wiki. No contexto geral, este procedimento leva a que o formulário criado possa ser utilizado para criar páginas, por exemplo, com o nome do livro “MediaWiki Book”, e que este livro fique também associado à categoria “BookCategory”.

A criação de uma *classe* através do *Semantic Forms* engloba todos os passos anteriormente descritos, com a diferença de que todos são introduzidos num só momento.

A Figura 3.5<sup>11</sup> mostra que são também alguns os projetos que utilizam esta extensão e tiram proveito das funcionalidades de inserção de dados através de formulários.

### 3.5 Resumo e conclusões

Este capítulo procurou descrever os conceitos chave relacionados com as wikis. Atualmente as wikis são utilizadas regularmente, nomeadamente em projetos de vertente colaborativa, como a Wikipédia. As wikis implementam diversas funcionalidades base e, para além disso, são de fácil aprendizagem para o utilizador.

O aparecimento da web semântica teve como objetivo estender a web atual e estruturá-la para a tornar mais funcional. Uma wiki semântica combina as wiki com a web semântica, o que permitiu a introdução de novas relações nos conteúdos de uma wiki.

Após uma análise detalhada de algumas wiki semânticas, a plataforma escolhida para o desenvolvimento do DataNotes foi a SMW e respetiva extensão *Semantic Forms*. Esta escolha assenta num conjunto de factos: este projeto é livre, existe um enorme conjunto de extensões que podem ser integradas e outras que podem ser desenvolvidas para integrar e por fim, existe também a possibilidade de melhoramentos em termos de armazenamento de dados.

<sup>11</sup>Disponível em: [http://www.mediawiki.org/wiki/Extension:SemanticForms/Sites\\_that\\_use\\_Semantic\\_Forms](http://www.mediawiki.org/wiki/Extension:SemanticForms/Sites_that_use_Semantic_Forms)



## Capítulo 4

# Anotação de conjuntos de dados apoiada numa wiki semântica

Após a tomada de decisão sobre o uso de uma wiki semântica procedeu-se a um estudo aprofundado dos requisitos do DataNotes. O DataNotes é uma plataforma colaborativa e permite aos investigadores anotar dados relativos a um projeto, com base em esquemas de metadados multidisciplinares previamente introduzidos na plataforma por um curador. Embora o DataNotes esteja concentrado na anotação dos dados e na preservação dos metadados, esta plataforma estabelece uma comunicação com a UPBox através de uma API. A UPBox é uma plataforma para depósito de dados no repositório da Universidade do Porto [Bar13]. De forma a integrar estas plataformas num processo de curadoria, os dados anotados no DataNotes estão relacionados com as pastas e ficheiros que foram depositados na UPBox.

### 4.1 Especificação de requisitos

Antes de iniciar a especificação de requisitos foi necessário reunir as ideias essenciais acerca do que o DataNotes iria disponibilizar aos investigadores com um conjunto de dados para anotar.

Após esta análise concluiu-se que, inicialmente, o investigador teria interesse em depositar esses dados num repositório e de seguida proceder à sua descrição através de uma aplicação web que permitisse uma interação intuitiva e colaborativa. A importância de trazer para junto de investigadores a oportunidade de anotar os dados produzidos por outros seria também interessante salvaguardando as devidas permissões. Para isso utilizou-se um conceito denominado por “Projeto”. Considera-se que a um projeto estão associados um conjunto de investigadores. Um projeto contém uma estrutura de pastas e ficheiros que devem ser anotados. Ainda que não seja o investigador A a inserir um ficheiro no repositório, caso pertença ao projeto (por atribuição de outro investigador) tem permissões para anotá-lo. A descrição dos dados iria permitir que se tornassem facilmente interpretáveis para os investigadores do projeto onde estão inseridos. Para elaborar as descrições dos dados seria de esperar que a aplicação sugerisse os esquemas de metadados ou normas, de forma a motivar o investigador para a descrição. Os esquemas de metadados que não

fossem sugeridos poderiam ser selecionados manualmente. Poderia também dar-se o caso de não existirem esquemas de metadados ou normas úteis para os dados a anotar, mas o investigador poderia optar por colocar qualquer descrição em texto livre.

Os itens seguintes mencionam os tópicos gerais que a extensão DataNotes deverá suportar.

- Permitir a introdução de esquemas de metadados e normas na wiki;
- No momento em que é criada uma anotação sugerir os esquemas de metadados ou normas registados na wiki;
- Armazenar toda a informação que é introduzida no momento do registo/edição de uma anotação;
- Gerar um ficheiro num formato RDF, que contém o modelo com descrições de um conjunto de dados da wiki;
- A longo prazo, manter toda a informação devidamente organizada, ou seja, tentar que os nomes atribuídos sejam coerentes com o que se está a fazer.

Estes itens gerais começam a aproximar-se daquilo que é pretendido como requisitos do sistema. Cada ator será responsável por um conjunto de tarefas, especificadas nas subsecções [4.1.1](#) e [4.1.2](#). Mais detalhes sobre os requisitos do DataNotes podem ser encontrados no anexo [A](#).

#### 4.1.1 Atores

Os atores ou utilizadores desta extensão wiki são essencialmente os **investigadores**, como já se tem vindo a referir, e os **curadores**. Os curadores têm em mão diferentes responsabilidades.

Atualmente, num processo de curadoria de dados, um curador tem o papel de intermediário ao longo de todo o processo. Apesar de se pretender que tenha uma presença menos frequente no desencadear do processo de anotações, continua a ter um papel indispensável. Será o curador o responsável pela introdução de normas e esquemas de metadados na plataforma.

Aos investigadores cabe anotarem os seus dados, e eventualmente os dados dos outros investigadores dos projetos aos quais pertencem através de uma interface simples e dinâmica.

#### 4.1.2 Casos de uso

O objetivo de mais alto nível é o desenvolvimento de um sistema de anotações de estruturas de diretórios baseado numa wiki semântica. Ao longo do processo de depósito e anotação de um conjunto de ficheiros ou apenas de um ficheiro, o(s) utilizador(es) têm à sua disposição um conjunto de funcionalidades. Tais funcionalidades podem ser acedidas através de um URL num *browser* em execução.

A Figura [4.1](#) ilustra o diagrama de casos de uso da extensão DataNotes com três partes distintas. Uma parte ilustra a gestão de utilizadores, outra a gestão de anotações e por fim a gestão

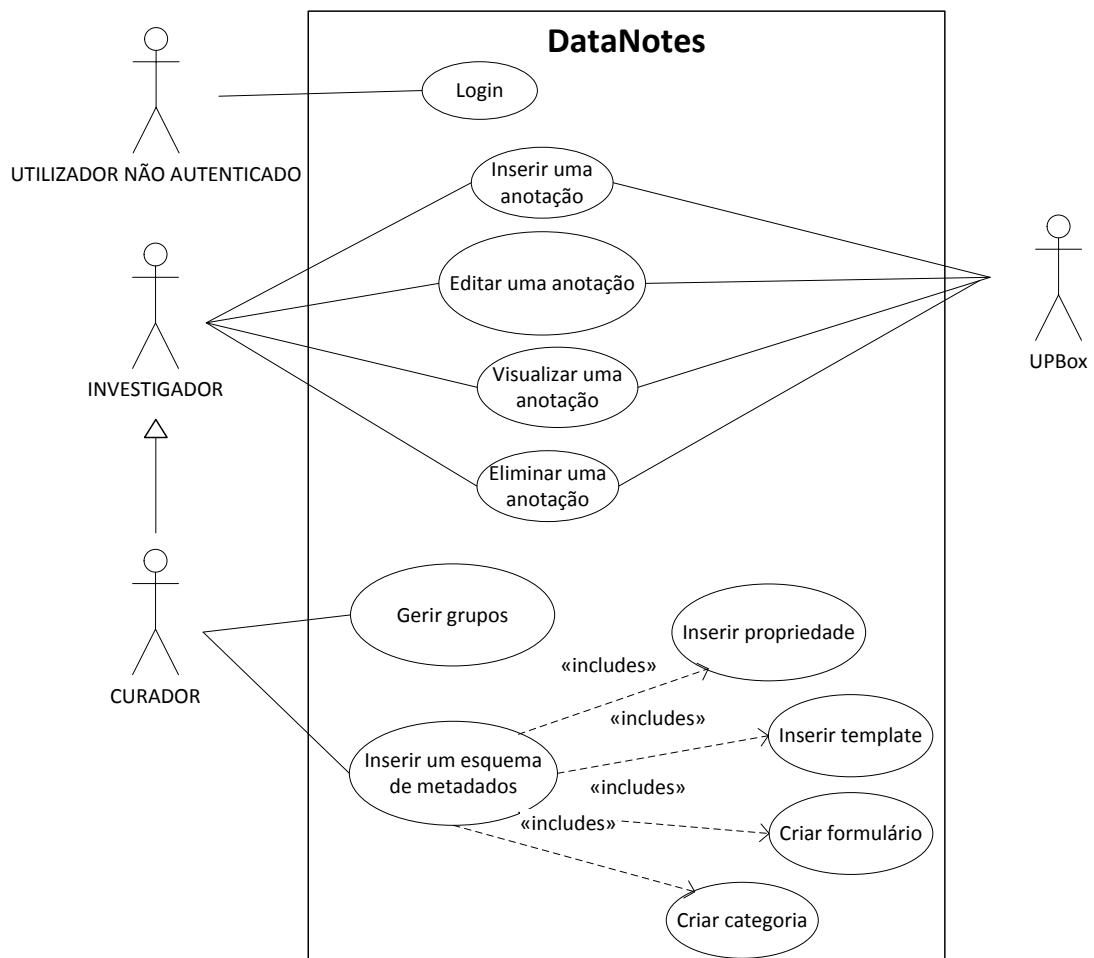


Figura 4.1: Casos de utilização do DataNotes

dos esquemas de metadados ou normas. De seguida, será exemplificado o fluxo de execução que se pretende em cada uma das partes:

- **Gestão de utilizadores** — Esta gestão, já definida pela *MediaWiki*, dá a oportunidade ao curador de registar novos utilizadores e atribuí-los a um ou mais grupos. Os grupos poderão ser criados pelo curador com as devidas permissões de acesso para os investigadores.
- **Gestão de anotações** — Após o *login* na wiki, o utilizador poderá seleccionar qual o projeto que pretende visualizar na sua estrutura de pastas e ficheiros. No decorrer desta ação irá surgir uma árvore com a estrutura referida onde o utilizador poderá indicar o que pretende anotar. Caso exista já uma anotação referente ao que foi selecionado, os metadados poderão ser visualizados. Caso contrário, uma nova página wiki surge para iniciar a anotação. Neste instante, alguns campos do esquema de metadados mais utilizado e conhecido, *Dublin Core*, são sugeridos para que o utilizador tenha uma pequena orientação. Não significa que o utilizador pretenda utilizar este esquema de metadados e se desejar pode mesmo eliminá-lo. No decorrer desta anotação existirá a possibilidade de escolher outros esquemas de metadados ativos na wiki e os respetivos descritores. No fim da página, existe uma caixa de texto livre onde o utilizador pode descrever aquilo que não foi possível descrever através dos esquemas de metadados. O objetivo é encorajar o utilizador a deixar uma descrição geral.
- **Gestão de esquemas de metadados ou normas** — Os esquemas de metadados nesta extensão fazem uso das funcionalidades de criação de propriedades, *templates*, formulários e categorias descritas na subsecção 3.4.1. O mesmo acontece quando se pretende introduzir qualquer outro tipo de esquemas de metadados, como o *Dublin Core*. Neste processo, os curadores terão de seguir todos os passos necessários. Sem, ou com poucos esquemas de metadados, um sistema de anotações torna-se desinteressante. É necessário investir na procura de esquemas de metadados e normas que cubram diversos domínios.

A extensão também deve obedecer a requisitos não funcionais, ou seja, deve ter algumas qualidades que embora não sejam visíveis para o utilizador são essenciais para que as funcionalidades a que este acede sejam executadas nas melhores condições:

- Manter todo o sistema o mais eficiente possível. A implementação de um sistema com qualidade permitirá que os seus utilizadores tenham uma plataforma bastante intuitiva para realizarem as devidas operações;
- A manutenção e a escalabilidade de todo o sistema;
- Uma vez que existe uma grande variedade de *browsers* no mercado, a extensão deve permitir que a portabilidade entre os vários tipos de *browser* seja possível.
- A usabilidade, principalmente da extensão, é um dos aspetos mais importantes a ter em consideração. Desta forma, a interface com o utilizador deve manter um aspeto minimalista para que o utilizador não tenha dificuldades em entender o que pode fazer nos diversos contextos com que se depara.

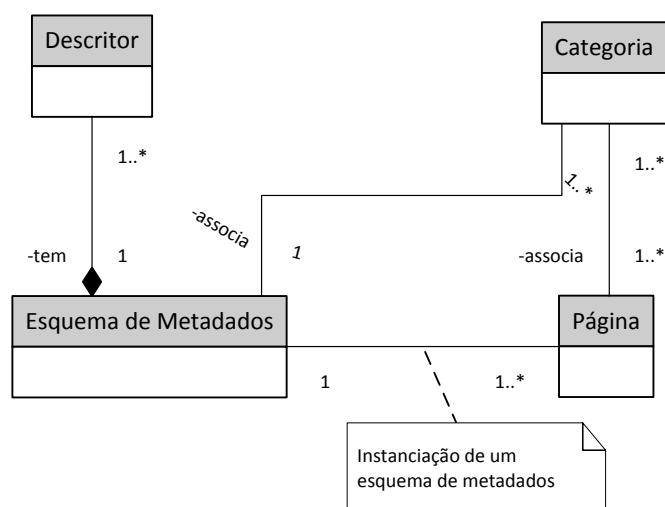


Figura 4.2: Modelo concetual do domínio antes da implementação da extensão DataNotes

## 4.2 Modelo concetual do domínio

O desenvolvimento desta extensão apoiada numa plataforma wiki, neste caso, na *MediaWiki* e *Semantic MediaWiki*, exigiu algumas adaptações face àquilo que já estava implementado. Os modelos concetuais do domínio ilustram simples diagramas que mostram o que foi necessário modificar para atingir os objetivos.

De forma a ilustrar o que se pode encontrar na presença do SMW e do *Semantic Forms*, a Figura 4.2 apresenta um modelo concetual do domínio antes da implementação do DataNotes, do qual se pode ler que:

- Um esquema de metadados por exemplo, *Dublin Core*, contém um conjunto de descritores que só existem dentro deste esquema;
- Uma página é uma instância de somente um esquema de metadados (naturalmente esta instância contém todos os descritores deste esquema de metadados);
- Uma página fica associada a uma ou várias categorias (por exemplo, sobre um *template* que caracteriza um país pode criar-se a categoria país e capital).
- Um esquema de metadados pode associar uma ou várias categorias.

Neste contexto, uma anotação faria referência a todos os descritores de um só esquema de metadados. Com esta perspetiva, os requisitos desta plataforma wiki como sistema de anotações não são cumpridos. É necessário que uma anotação possa fazer referência a descritores de um ou mais esquemas de metadados.

Assim, alterou-se o modelo concetual do domínio, de forma a obter as restrições encontradas na Figura 4.3. Aqui podemos observar que:

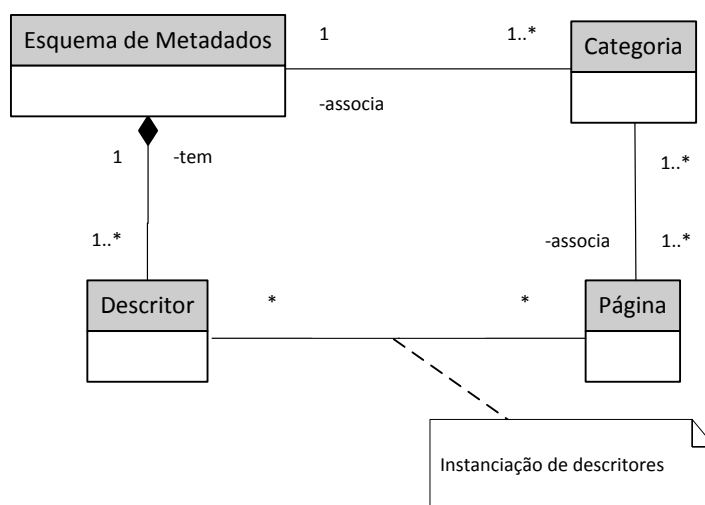


Figura 4.3: Modelo conceitual do domínio depois da implementação da extensão DataNotes

- Um esquema de metadados por exemplo, *Dublin Core*, contém um conjunto de descritores que só existem dentro deste esquema;
- Uma página está associada a um ou vários descritores de um ou vários esquemas de metadados, o que faz com que uma página seja uma instância de um conjunto de descritores. Por exemplo, podem ser escolhidos dois descritores do esquema de metadados *Dublin Core* e somente um descritor de outro esquema de metadados;
- Uma página fica associada a uma ou várias categorias;
- Um esquema de metadados pode associar uma ou várias categorias.

Note-se que a SMW e o *Semantic Forms* estão desenhados para o modelo apresentado inicialmente (Figura 4.2) e não para aquilo que se pretende implementar (Figura 4.3). Toda a estrutura que assenta sobre a ideia inicial é sólida e difícil de modificar. Porém, este desafio encorajou a procura de uma solução viável para a resolução do problema que será apresentada no Capítulo 5.

### 4.3 Resumo e conclusões

O estudo e a identificação dos requisitos e funcionalidades que a extensão requer são fundamentais para o decorrer do desenvolvimento de qualquer projeto. O objetivo desta extensão focou-se no desenvolvimento de um sistema colaborativo de anotação de estruturas de diretórios destinado ao uso por parte dos investigadores. Esta extensão foi desenvolvida sobre a plataforma MW e SMW e com recurso a uma extensão *Semantic Forms*. O DataNotes deverá suportar esquemas de metadados de diversos domínios introduzidos manualmente pelos curadores, para que os investigadores possam utilizar descritores de alguns dos esquemas de metadados para anotar os seus dados.



## Anotação de conjuntos de dados apoiada numa wiki semântica

Numa fase inicial, a extensão irá necessitar de ser abastecida com alguns esquemas de diversos domínios. O crescimento desta extensão exige a colaboração contínua de investigadores a anotar os seus dados e dos curadores a introduzir novos esquemas de metadados ou normas na plataforma.



## Capítulo 5

# Arquitetura da extensão DataNotes

Neste capítulo é apresentada a arquitetura da extensão DataNotes. Começa-se por realçar a visão geral e os componentes necessários para a sua implementação, bem como o modo que os componentes se relacionam. De seguida, cada um desses componentes é descrito com maior detalhe e são identificadas as suas principais funcionalidades.

### 5.1 Visão Geral

Como foi referido anteriormente, o objetivo deste trabalho é desenvolver uma extensão à plataforma *Semantic MediaWiki*. A arquitetura do sistema inclui a arquitetura lógica (camadas/módulos), a arquitetura física (nós, componentes e ligações) e a arquitetura tecnológica (tecnologias, linguagens e plataformas).

A arquitetura da SMW apresentada na Figura 5.1 divide-se em componentes e cada um tem código específico baseado num modelo de dados comum que proporciona uma apresentação unificada de dados semânticos. A MW contém cinco componentes distintos apoiados num servidor *web Apache* e sobre cada um deles assentam outros componentes da SMW.

O componente de manipulação e visualização de páginas na MW é responsável por mostrar os dados inseridos nas suas páginas. Dentro deste componente existem classes responsáveis por compreender a linguagem wiki base e por processá-la para outro formato, como por exemplo para HTML. A SMW introduz novas sintaxes de linguagem wiki e, por este motivo, também possui algumas classes para a compreensão e o processamento desta nova sintaxe. As classes de compreensão e o processamento comunicam com a base de dados da MW e com as tabelas que se estendem para a SMW, de forma a dar suporte ao armazenamento e preservação dos conteúdos.

O componente de páginas especiais da MW recebe todas as páginas criadas por extensões da SMW com diversos tipos de dados. Como resultado do processamento pode haver comunicação com o componente base de manipulação e visualização de páginas, bem como com a base de dados de ambas as partes. Por acréscimo é possível criar diferentes estilos a adotar para a apresentação das páginas, bem como interações dinâmicas entre os componentes das páginas. Estes ficheiros *.css* e *.js* não necessitam de ser processados por nenhum componente da MW.

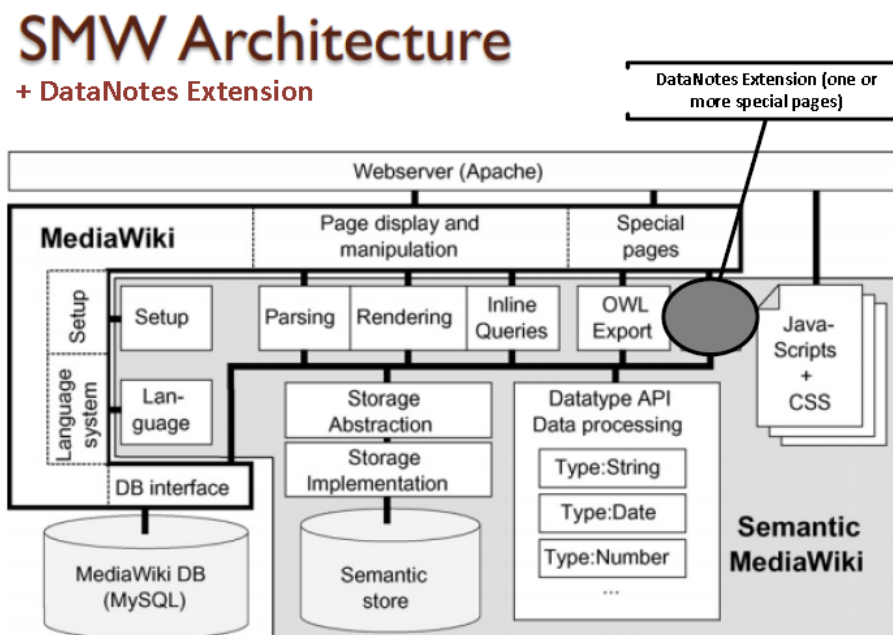


Figura 5.1: Arquitetura SMW (in [KVV<sup>+</sup>07]) com a extensão DataNotes

Algo importante que se pode retirar da Figura 5.1 é que todas as extensões que vierem por acréscimo a esta plataforma serão forçadas a seguir o protocolo de integração já estudado. Isto acontece porque, apesar de serem implementadas novas funcionalidades, é sempre a camada mais baixa, neste caso a MW, que é responsável por desenhar as páginas. Neste contexto, o DataNotes é um conjunto de páginas especiais que pertence à plataforma SMW.

## 5.2 Componentes

Com base naquilo que foi explicado anteriormente e apesar do DataNotes na arquitetura da SMW não ser mais do que um conjunto de “páginas especiais”, a Figura 5.2 mostra um diagrama de componentes elaborado para aquilo que se pretende implementar com o DataNotes.

A leitura deste diagrama é efetuada na vertical, de cima para baixo. Em primeiro lugar, os dois serviços são responsáveis pela comunicação com o utilizador ou com outra máquina. Ambos os serviços mostram os conteúdos de entrada e saída para os componentes do DataNotes.

- **Interação com a UPBox** — A UPBox é uma plataforma de armazenamento e partilha na nuvem de ficheiros e pastas para a comunidade de investigadores. Funciona como uma pasta partilhada numa rede local via *web*. O uso deste serviço permite fazer pedidos através de dois URL. O primeiro retorna informação sobre os projetos de um investigador. O segundo, para cada projeto, devolve a respetiva estrutura de diretórios e ficheiros.

## Arquitetura da extensão DataNotes

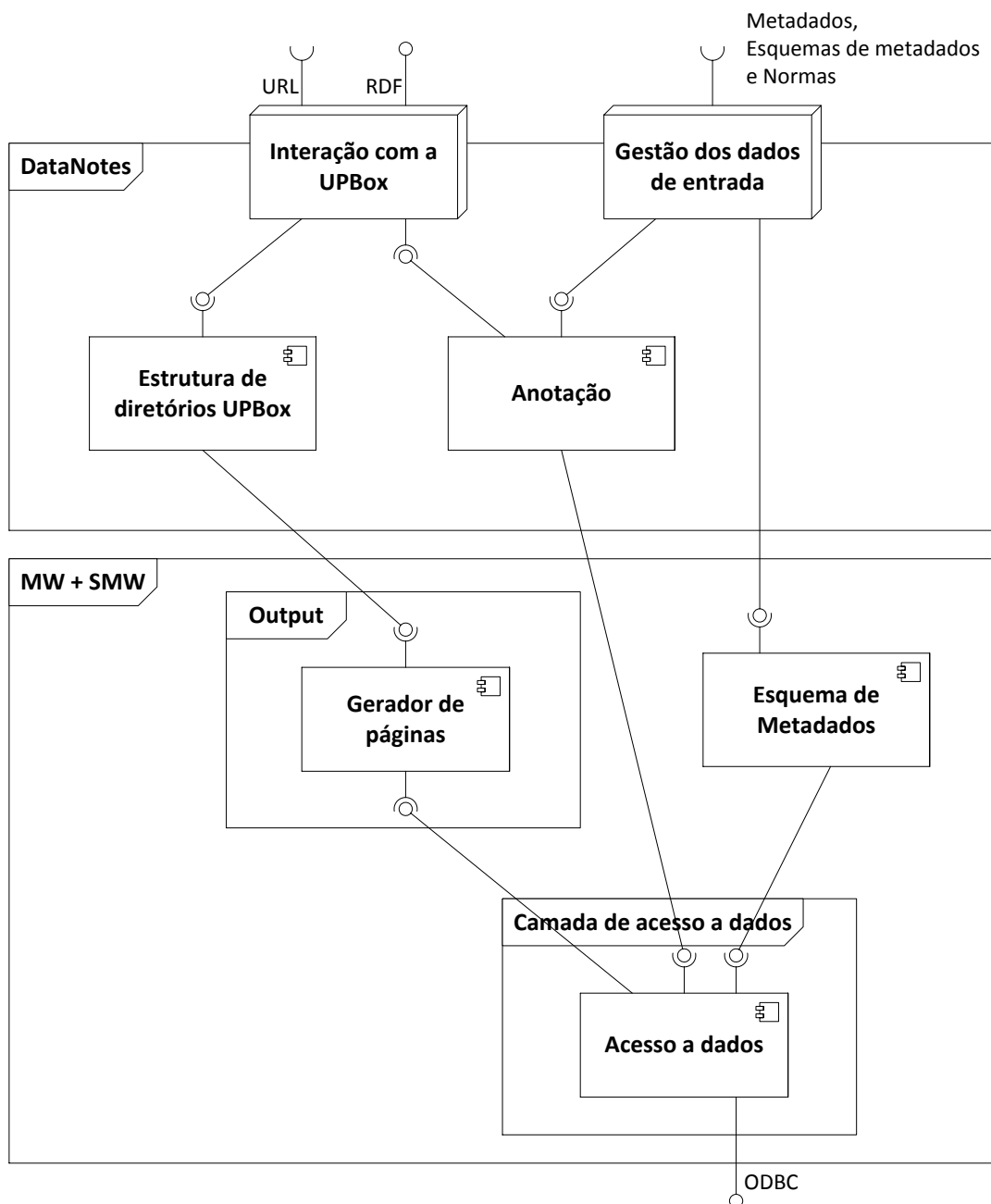


Figura 5.2: Diagrama de Componentes do DataNotes

- **Gestão dos dados de entrada** - Aqui existe a entrada de dados por parte de um utilizador. Estes dados podem ser esquemas de metadados ou normas, bem como os metadados do objeto a anotar.

De seguida, são descritos os componentes deste diagrama.

- **Estrutura de diretórios UPBox** — Feito o pedido ao URL que devolve o nome dos projetos, a classe responsável pela manipulação da informação em formato JSON recebida, preenche uma *combobox* e envia a informação para o componente “Gerador de páginas”. Assim que algum projeto é selecionado por um utilizador, é feito um pedido ao segundo URL que devolve a estrutura de diretórios e ficheiros referente a esse projeto também em formato JSON. Recebida e processada a informação é enviada para o componente “Gerador de páginas”.
- **Anotação** — Este componente é responsável por gerar as páginas relativas às anotações. É necessário que durante a escolha dos esquemas de metadados e respetivos descritores que sejam também introduzidas descrições para esses campos. O código HTML é convertido em código de linguagem wiki, que por sua vez é enviado para o componente de “Acesso a dados” para ser gravado na base de dados.
- **Esquema de Metadados** — No contexto do DataNotes um esquema de metadados é um conjunto de páginas disponibilizadas pelo *Semantic Forms*. Estas páginas permitem a criação de propriedades, *templates* (ou seja, a criação de descritores e respetivas propriedades semânticas), formulários de acordo com os *templates* disponibilizados e por fim categorias. Em caso alternativo pode optar-se pela criação de uma classe que engloba todas as fases anteriores. Este código está disponível no módulo da MW e da SMW onde o DataNotes o utiliza para a criação dos seus esquemas de metadados.
- **Acesso a dados** — Este componente recebe dados vindos de outros componentes e armazena-os na base de dados da MW e SMW. A informação é guardada em linguagem wiki recorrendo ao conetor ODBC disponibilizado pelo MySQL.
- **Gerador de páginas** — O gerador de páginas tem um papel fundamental em todo este conjunto de componentes. Recebe os conteúdos, essencialmente vindas do componente “Acesso a dados”, para que sejam processadas as páginas wiki.

### 5.3 Tecnologias

Ao longo do desenvolvimento do projeto e consequente implementação, foram utilizadas diversas tecnologias como é ilustrado na Figura 5.3.

O *Apache HTTP Server* é um servidor web gratuito escolhido por inúmeros projetos que envolvem tecnologias de transmissão via web, processamento de dados e execução de aplicações distribuídas. Na prática, o PHP utilizado pela plataforma MW e SMW complementa o *Apache*

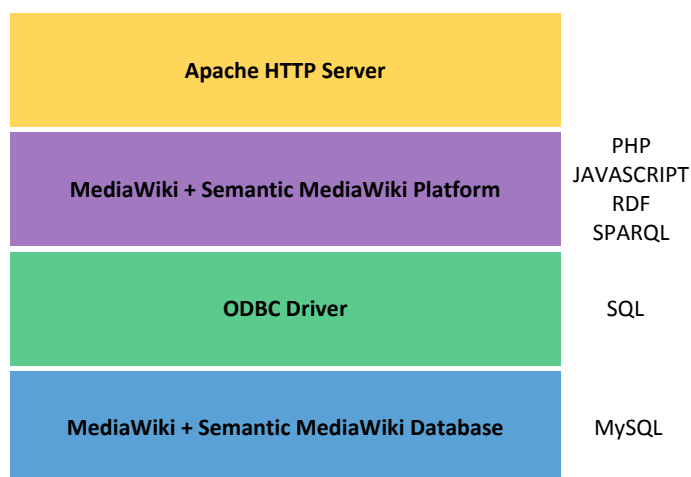


Figura 5.3: Pilha de tecnologias

*HTTP Server* e processa os pedidos de determinados tipos de ficheiros (tipicamente os que terminam em .php). Adicionalmente, através do conetor ODBC são disponibilizados um conjunto de métodos para aceder à base de dados. A opção da utilização da MW e SMW levou a que a base de dados estivesse já implementada com o recurso ao sistema de gestão de base de dados MySQL que faz uso da linguagem SQL.

## 5.4 Metodologias

Existem diversas metodologias de desenvolvimento de software que têm como finalidade definir um método de trabalho para a correta implementação do sistema. Optou-se por adotar um processo iterativo que consiste na realização de um conjunto de etapas devidamente organizadas, como ilustra a Figura 5.4.

Após a definição da arquitetura e a escolha das novas funcionalidades a implementar, o trabalho foi dividido em iterações, consistindo cada uma delas num ciclo que inclui a implementação de uma funcionalidade, seguida da adição de conteúdos de teste, para finalmente a testar. Este ciclo será mantido até que a funcionalidade esteja concluída sem erros, pelo que é possível afirmar que este processo assenta numa prototipagem evolutiva.

## 5.5 Resumo e conclusões

Uma vez escolhida a plataforma MW e SMW para a implementação do DataNotes foi necessário elaborar uma análise da arquitetura existente e compreender onde iria integrar a nova extensão. Este estudo permitiu a elaboração de um diagrama de componentes para ilustrar a interação e integração do DataNotes com os outros componentes. Através da pilha de tecnologias pretendeu-se mostrar quais seriam as necessidades de cada camada de implementação.

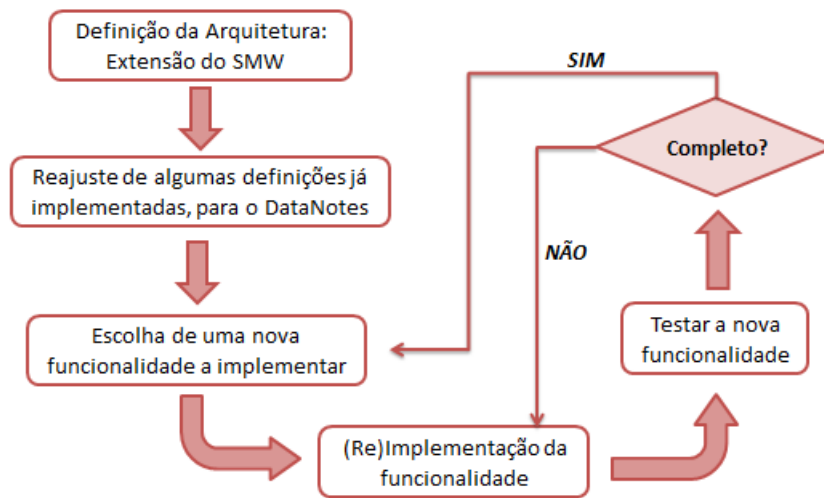


Figura 5.4: Metodologia de implementação

Por fim ilustrou-se, através de um diagrama, as metodologias adotadas para o desenvolvimento da extensão.



## Capítulo 6

# Implementação do DataNotes

Este capítulo tem como objetivo documentar a implementação do DataNotes, através de uma descrição detalhada, e apresentar as decisões que foram tomadas.

Quando o desenvolvimento de um novo projeto tem por base outro previamente implementado é necessário adquirir inicialmente alguns conhecimentos. Em projetos *open-source* a documentação disponível serve para perceber o que foi elaborado até à data e eventualmente como se podem estender novas funcionalidades. Contudo, quando o projeto tem uma dimensão considerável a aprendizagem é contínua.

### 6.1 Base de dados da MediaWiki e Semantic MediaWiki

Toda a informação que é colocada na wiki é registada na base de dados. Atualmente, num total de cinquenta e três tabelas que constituem esta base de dados, as cinco apresentadas na Figura 6.1<sup>1</sup> são as responsáveis pelo armazenamento da informação principal.

A tabela *Page* é o núcleo da wiki. Cada entrada nesta tabela corresponde ao título da página criada e respetiva metainformação. No entanto, o texto da página é armazenado na tabela *Text*. Deste modo, o atributo “page\_latest” (chave estrangeira) é utilizado para pesquisar na tabela *Revision* o “rev\_id” (chave primária). Assim consegue obter-se o “rev\_text\_id” (chave estrangeira) que é utilizado para procurar o “old\_id” (chave primária) na tabela *Text* e finalmente obter o texto associado. Quando uma página da wiki é editada, é também adicionada uma nova entrada à tabela *Revision*. Esta tabela é muito semelhante à *Recentchanges*. A primeira é utilizada para manter o histórico da página e listagens de contribuições do utilizador e pode ser mantida durante tempo indefinido. A segunda é usada para registar as mudanças recentes, as listas de vigilância e, no caso de criação de uma página, a lista de novas páginas. Estes itens são periodicamente eliminados. Por fim, a tabela *Pagelinks* regista todas as ligações internas na wiki. Além destas cinco tabelas base existem outras tabelas que pertencem a esta plataforma. É o caso da tabela que regista as categorias que se destinam a agrupar páginas sobre temas semelhantes, que leva os utilizadores a ter uma interação simplificada na navegação, classificação e localização de artigos relacionados.

---

<sup>1</sup>Disponível em: [http://www.mediawiki.org/wiki/Manual:Database\\_layout](http://www.mediawiki.org/wiki/Manual:Database_layout)

## Implementação do DataNotes

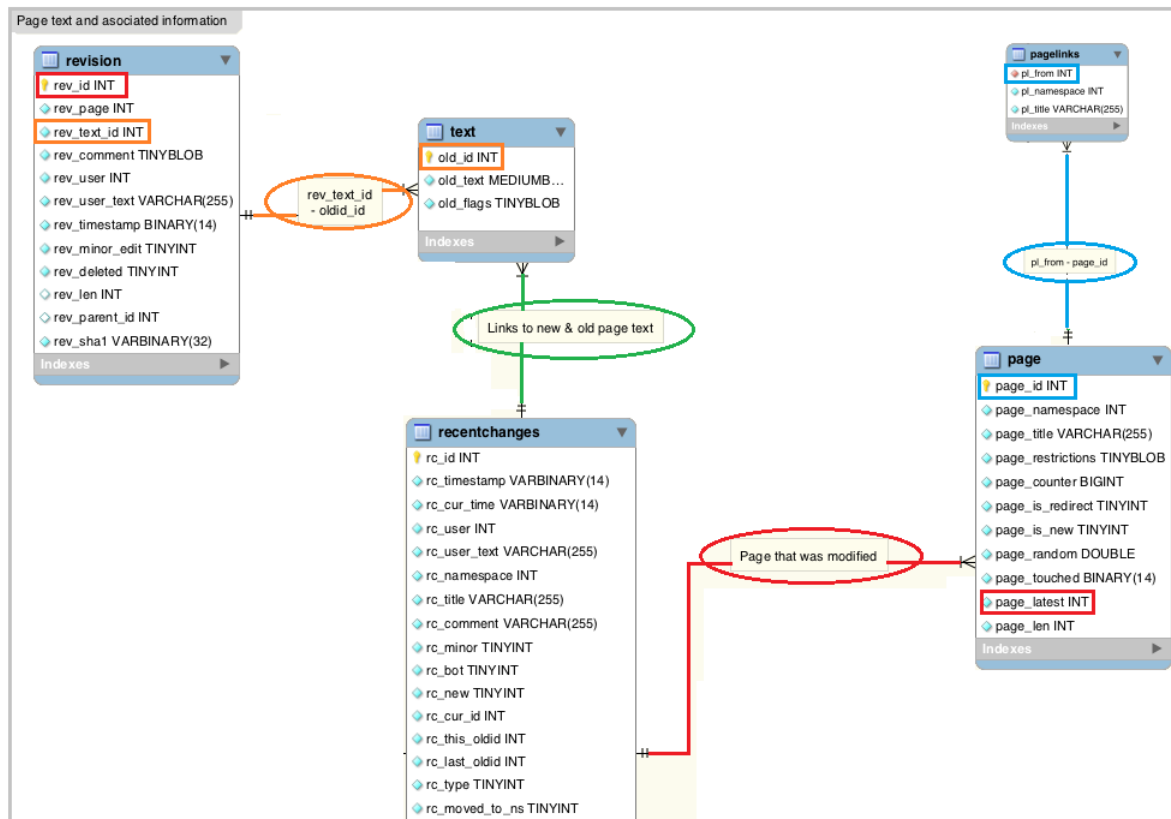


Figura 6.1: Núcleo da base de dados da MediaWiki

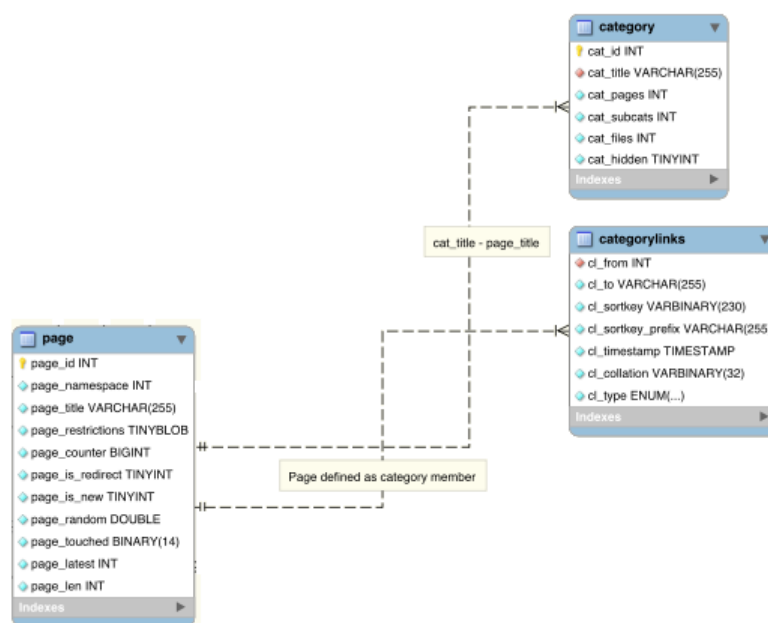


Figura 6.2: Categorias na base de dados da MediaWiki

A Figura 6.2 <sup>2</sup> representa a tabela *Page* com a respetiva ligação às tabelas de categorias. A tabela *Category* regista os temas introduzidos na wiki e a tabela *Categorylink* regista a terminologia utilizada nas páginas que pertencem a uma categoria, como por exemplo, `[[Category:Title]]`.

## 6.2 Ficheiros base da MediaWiki e Semantic MediaWiki

Os conjuntos de ficheiros PHP da MW disponibilizam funções para o acesso à base de dados, bem como para a criação de páginas, categorias, entre outros. Os dois ficheiros mais conhecidas e importantes são o *Article.php* e o *Title.php*.

- *Article.php* — Possui funções que permitem a criação de uma nova página sobre um artigo na wiki e o acesso ao seu conteúdo caso já exista;
- *Title.php* — Representa os títulos que existem nas páginas MW, quer estes representem artigos, *templates*, propriedades ou categorias. Esta classe permite verificar se um determinado título já existe, bem como criar um novo no caso contrário.

Da mesma forma, os conjuntos de ficheiros PHP da extensão SMW disponibilizam também as suas funções:

- *SMW\_PropertyPage.php* — O objetivo desta classe é apresentar os conteúdos relativos a uma propriedade criada pelo utilizador, bem como as páginas que utilizam tal propriedade;
- *SMW\_SQLStore2.php* — Esta classe implementa métodos que permitem o acesso à base de dados MW;
- *SMW\_SemanticData.php* — É através desta classe que são representados os dados semânticos de um determinado artigo numa página wiki.

O ficheiro de configuração *LocalSettings.php* fornece configurações locais da instalação da MW, desde o endereço onde está instalado, à informação de criação e configuração da base de dados. Depois da sua instalação, este ficheiro é gerado automaticamente, embora seja comum ajustar alguns dos parâmetros ou adicionar novos. Alterar uma configuração geralmente significa mudar o valor de uma variável PHP.

## 6.3 Criação de extensões na MediaWiki e Semantic MediaWiki

As extensões que podem ser criadas sobre a MW apresentam naturezas distintas <sup>3</sup> de acordo com aquilo que se pretende implementar:

- *Tag Extensions* — Permite aos utilizadores criar novas tags personalizadas;

---

<sup>2</sup>Disponível em: [http://www.mediawiki.org/wiki/Category:MediaWiki\\_database\\_tables](http://www.mediawiki.org/wiki/Category:MediaWiki_database_tables)

<sup>3</sup><http://www.mediawiki.org/wiki/Manual:Extensions>

- *Parser Functions* — Integram com o *parser* geral e analisam as novas tags introduzidas;
- *Skins* — Permitem que os utilizadores personalizem a aparência da MW;
- *Magic Words* — É uma técnica de mapeamento da linguagem wiki para um único identificador que está associado a uma função;
- *Special Pages* — Descrito na sub-secção 6.3.1;
- *Hooks* — Descrito na sub-secção 6.3.2.

### 6.3.1 Páginas especiais

Para o DataNotes, a natureza escolhida foi a implementação de páginas especiais<sup>4</sup>. As páginas especiais consistem na criação de uma página wiki onde as funcionalidades criadas são integradas na wiki. Uma extensão com esta natureza é composta por ficheiros PHP elaborados de acordo com as especificações pretendidas pela MW. As normas definidas descrevem a criação básica de quatro ficheiros, sendo eles:

**<special\_page\_name>.php** — Ficheiro de instalação. É o local onde são definidas as localizações dos outros ficheiros (classes, mensagens, hooks e folhas de estilo CSS), descrição da extensão, entre outros.

**<special\_page\_name>.body.php** — Ficheiro onde são criadas as funcionalidades que a extensão permite serem executadas como parte integrante da wiki. Podem existir outros ficheiros que complementem este, eventualmente outras classes que sejam criadas como auxiliares.

**<special\_page\_name>.i18n.php** — Ficheiro que permite definir mensagens e textos para as diversas línguas permitidas na MW. No entanto, o utilizador poderá solicitar que os conteúdos visualizados estejam somente na língua definida durante a instalação da MediaWiki.

**<special\_page\_name>.alias.php** — Ficheiro que tem como objetivo a tradução do título das extensões. Isto permite que o acesso a um URL possa ser feito tanto *.../Special:ExportRDF* como *.../Especial:Exportar\_RDF*.

### 6.3.2 Hooks

A natureza escolhida para a integração da estrutura de diretórios vinda da UPBox tem por base a utilização de *hooks*<sup>5</sup>. Tal como a tradução à letra menciona, *hooks* são ganchos que permitem que determinado código seja executado num evento. Para isso é necessário colocar no código do evento, o local exato onde o gancho deve ser executado.

- **Definição do Hook:** `$wgHooks['EventName'][] = 'MyExtensionHooks::someFunction';`
- **Definição da função:** `public static function someFunction( parameters ) ... ;`
- **Chamada do Hook:** `wfRunHooks( 'EventName', parameters ).`

<sup>4</sup>[http://www.mediawiki.org/wiki/Manual:Special\\_pages](http://www.mediawiki.org/wiki/Manual:Special_pages)

<sup>5</sup><http://www.mediawiki.org/wiki/Manual:HooksManual>

Ao longo de todo o código desenvolvido para a MW e SMW existem chamadas a diversos hooks <sup>6</sup>. Isto facilita o desenvolvimento das novas extensões para evitar interferir em código previamente implementado.

### 6.3.3 Ficheiro de configuração

Tal como referido anteriormente é necessário acrescentar conteúdos ao ficheiro *LocalSettings.php* através da indicação de que foram adicionadas novas extensões, bem como alterações nas permissões dos utilizadores:

- **Incluir novas extensões:** `include_once("$IP/extensions/SemanticForms/SemanticForms.php");`
- **Alterar permissões de grupos:** `$wgGroupPermissions['Researchers']['edit'] = true;`

## 6.4 Extensão DataNotes

A aprendizagem das noções base para a criação de ficheiros na MW e na SMW tornou-se fundamental para prosseguir e planear como deveria desencadear-se a implementação do DataNotes.

Após a instalação das plataformas necessárias, ou seja, MW e SMW com a respetiva extensão *Semantic Forms*, as atenções desviaram-se para a vertente de experimentação e interpretação do processamento da informação dentro desta wiki, desde a forma como eram inseridos e manipulados os dados à gravação na base de dados. A inserção de texto na wiki em linguagem natural é processada por código e convertida para linguagem wiki. Como também já foi possível apurar, a base de dados da MW e SMW, para a quantidade de conteúdos que armazena, não contém muitas tabelas. Embora isto leve a custos elevados no processamento de código que está armazenado na base de dados e vice versa, de outra forma não seria possível esta wiki abranger tantas áreas. Normalmente, quando um projeto é implementado para registar os batimentos cardíacos de um ser humano sabe-se exatamente quais são os dados de entrada necessários. Se existir intenção de utilizar este projeto aplicado ao registo de intensidade do sopro de um ser humano, já não é adaptável a não ser que se modifiquem as tabelas da base de dados. A MW e SMW foram criadas não a pensar num projeto específico mas que fosse abrangente a tudo aquilo que se pretendesse utilizar, seja no domínio da engenharia hidráulica ou alimentar.

### 6.4.1 Introdução de esquemas de metadados na wiki

Face à realidade encontrada na introdução do esquema de metadados *Book* na subsecção 3.4.1, ou seja, a possibilidade de introdução de propriedades, um conjunto de descritores que usam essas propriedades (*templates*), a criação de formulários baseado nesses *templates* e a criação de categorias associadas aos formulários, concluiu-se que esta abordagem poderia ser utilizada para o contexto do DataNotes. Além disso, existe também a possibilidade de se inserir diretamente o conjunto dos componentes, através da criação de classes no *Semantic Forms*. Com base em

---

<sup>6</sup>[http://www.mediawiki.org/wiki/Manual:Hooks#Hooks\\_grouped\\_by\\_function](http://www.mediawiki.org/wiki/Manual:Hooks#Hooks_grouped_by_function)

todas estas funcionalidades seria perfeitamente aceitável criar uma ponte para um dos princípios do DataNotes: a criação de esquemas de metadados.

Através destas funcionalidades, juntamente com a possibilidade de introdução de dados via formulários (caixas de texto), torna-se simples que o curador possa criar os esquemas de metadados na wiki sem ter que se preocupar com qualquer tipo de sintaxe exigida. O processo pode ser inicialmente confuso mas, acaba por se tornar como outra qualquer ação.

### 6.4.2 Criação de formulários com base em esquemas de metadados

Posteriormente à introdução de esquemas de metadados na wiki, o passo seguinte deveria enquadrar a tentativa de introdução de dados na wiki com base nestes esquemas. Após o momento de experimentação facilmente se conclui que apenas se podia utilizar um único formulário para a introdução de dados. Tal como visto anteriormente com os modelos conceituais do domínio, isto não era o pretendido e aqui entra a necessidade de criação de uma nova extensão como o DataNotes e reestruturação de alguns conceitos face ao existente.

No decorrer da procura de alternativas para solucionar este desafio surgiu a ideia que acabou por ser distinguida. Abstraindo da forma como estavam implementadas as funcionalidades do *Semantic Forms* surgiram as seguintes necessidades:

#### Criar uma anotação

1. Um método que devolva todos os formulários e respetivos *templates* que existem na base de dados;
2. Reaproveitar o processamento que o *Semantic Forms* faz ao único *template* que vem da base de dados, acabando por o transformar em código HTML;
3. Colocar este código HTML numa tag `<div>` escondida;
4. Desenvolvimento do código e interação com o utilizador em linguagem *JavaScript*.
5. Reajustar o código existente para a gravação dos dados introduzidos sobre os esquemas de metadados, seguindo a lógica aplicada à gravação dos dados no *Semantic Forms*.

O item 1 permitirá saber todos os esquemas de metadados que estão gravados na wiki. Um exemplo de um resultado devolvido para um esquema existente é ilustrado na Figura 6.3.

Contudo é necessário manipular esta linguagem wiki para um formato HTML. O item 2 é responsável por essa transformação independentemente do número de esquemas de metadados que existam. O objetivo do item 3 é auxiliar aquilo que se pretende fazer no item 4. A ideia baseia-se na criação de uma página dinâmica recorrendo ao uso do *JavaScript*, onde inicialmente são disponibilizados apenas os três descritores mais comuns do *Dublin Core* (assume-me que o *Dublin Core* é o esquema de metadados base referência a ser usado num sistema com estas características). Os restantes esquemas e descritores podem ser selecionados pelos utilizadores ou

## Implementação do DataNotes

```
{{DC
|Title=
|Subject=
|Description=
|Type=
|Source=
|Relation=
|Coverage=
|Creator=
|Publisher=
|Contributor=
|Rights=
|Date=
|Format=
|Identifier=
|Language=
}}
```

Figura 6.3: Esquema de metadados Dublin Core

também podem ser retirados se houver algum engano. Este dinamismo é apoiado pelo código que está escondido na tag `div` HTML.

De forma a assegurar este procedimento foram desenvolvidos diversos métodos, entre eles os que permitiam obter os nomes dos esquemas de metadados, os nomes dos descritores de um esquema selecionado e se um determinado esquema já estava a ser visualizado. Esta interação com o utilizador será detalhada no Capítulo 7.

Para apoiar esta implementação foi necessária a criação de páginas especiais dentro da extensão DataNotes. Apesar do DataNotes juntamente com a UPBox trazer uma maior abrangência às fases de preservação e curadoria de dados, sendo este um projeto desenvolvido como um só, é também possível descrever dados que não pertencem à UPBox. Na Figura 6.4 é possível visualizar a classe da página especial *DN\_Annotate.php* que é uma ponte para a classe da página especial *DN\_AnnotateForm.php*. Caso se pretenda anotar outros dados que não tenham ligação com a UPBox poderá optar-se por selecionar na opção “Annotate” como ilustra a Figura 6.5 e introduzir um nome à anotação que se pretende fazer. Depois será reencaminhado para a página especial *DN\_AnnotateForm.php* que permite introduzir as descrições dos dados, tal como se tivesse optado pela anotação de dados de ficheiros vindos da UPBox.

### Editar uma anotação

A classe *DN\_FormEdit.php* da Figura 6.4 implementa a edição de uma anotação que segue a abordagem utilizada para a criação de uma anotação. Os itens anteriormente numerados para a criação de uma anotação mantêm-se ainda que com diferenças no código que é gerado. Ou seja, junto do código HTML escondido na tag `<div>` HTML podem encontrar-se não só os esquemas de metadados e os descritores como também algumas descrições anteriormente adicionadas que necessitam agora de ser mostradas. Quando o utilizador abre a página para anotação, o conteúdo

## Implementação do DataNotes

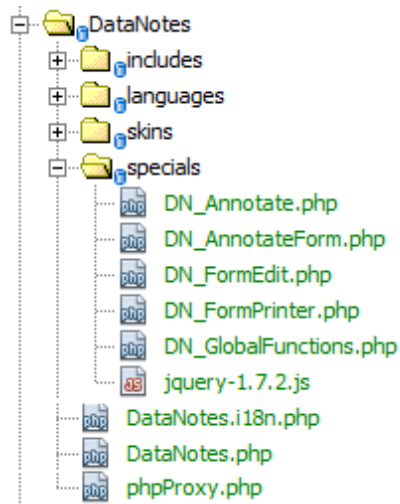


Figura 6.4: Classes das páginas especiais da extensão DataNotes

### Semantic MediaWiki

- Admin functions for Semantic MediaWiki
- Browse wiki
- Export pages to RDF
- Query Creator
- Search by property
- Semantic search

### Semantic Forms

- Create a category
- Create a class
- Create a form
- Create a property
- Create a template
- Edit with form
- Run query
- Start of form

### Semantic Annotations

- Annotate

Figura 6.5: Páginas especiais na wiki com a extensão DataNotes



mostrado é aquele que já tinha sido preenchido anteriormente. De seguida, a página pode ser manipulada tal como foi explicado para a criação.

### Visualizar uma anotação

Após a gravação dos dados, no contexto do *Semantic Forms*, os descritores dos formulários não usados eram mostrados, por definição, como campos vazios. No entanto, se os descritores não são escolhidos naquele esquema de metadados é porque não existe interesse que estejam contidos numa determinada descrição. Na procura de uma solução que manipulasse esta ideia, apoiada pela visualização do *Semantic Forms*, surgiu um *hook*.

- `$wgHooks['OutputPageBeforeHTML'][] = 'DN_AnnotateForm::onOutputPageBeforeHTML';`

Este item mostra a declaração do hook “*OutputPageBeforeHTML*” que permite a manipulação do conteúdo de uma página antes da sua visualização. Na classe *DN\_AnnotateForm* foi declarado e implementado o método *onOutputPageBeforeHTML* capaz de remover as linhas vazias do conteúdo da página recebida (neste caso o conteúdo é uma tabela), tal como se pode visualizar na Figura 6.6.

### 6.4.3 Criação de uma árvore com estruturas de ficheiros e diretórios da UPBox

Sendo este projeto destinado à anotação de pastas e ficheiros e estando integrado num processo de curadoria e preservação de dados, a sua relação com um outro projeto que permita o armazenamento de dados, como a UPBox, leva a que ambos os projetos se tornem como um só. Com vista a integrar estas duas plataformas tornou-se necessário perceber todo o fluxo de trabalhos que seria expectável para a sua interligação.

Em primeiro lugar surgiram questões relativas à autenticação. Um investigador tende naturalmente a autenticar-se na UPBox e após o depósito de algum ficheiro ou até a criação de um novo diretório poderá aceder ao DataNotes para anotar. A existência de uma nova autenticação no DataNotes tornar-se-ia inoportuno para o utilizador. Como solução, o envio de um sinal codificado por parte da UPBox seria suficiente. Mas, sempre que o utilizador pretendesse aceder ao DataNotes sem ter que passar pela UPBox, faria sentido que o seu nome de utilizador fosse o mesmo utilizado no contexto da UPBox e até no contexto utilizado pela Universidade do Porto no acesso ao SIGARRA (Sistema de Informação para Gestão Agregada dos Recursos e dos Registos Académicos). A metodologia mais correta seria o DataNotes integrar a autenticação por LDAP (Lightweight Directory Access Protocol). Existe uma extensão para a MW que implementa este tipo de autenticação <sup>7</sup>, mas tirar proveito das suas funcionalidades não seria intuitivo dado que a extensão não é facilmente integrável. Assim optou-se pela utilização da autenticação que vem já integrada com a plataforma MW.

Posteriormente também se colocou o caso de na hierarquia de pastas adotar a seguinte abordagem: após a anotação de uma determinada pasta, todas as sub-pastas e respetivos ficheiros não

<sup>7</sup>[http://www.mediawiki.org/wiki/Extension:LDAP\\_Authentication](http://www.mediawiki.org/wiki/Extension:LDAP_Authentication)

## Implementação do DataNotes

Title	A experiência n.º1 GRAVEMIT
Creator	Maria Leonor
Subject	
Description	A experiência decorreu durante a aula n.º2 da cadeira GRAV.
Publisher	
Contributor	
Date	2012/10/06
Type	
Format	DataSet
Identifier	
Source	
Language	
Relation	
Coverage	
Rights	



Title	A experiência n.º1 GRAVEMIT
Creator	Maria Leonor
Description	A experiência decorreu durante a aula n.º2 da cadeira GRAV.
Date	2012/10/06
Format	DataSet

Figura 6.6: Página de visualização — O antes e o depois

anotados herdavam a anotação da pasta mãe. Isto pode não significar que a anotação na pasta mãe partilhe todas as suas descrições às sub-pastas e ficheiros. Da mesma forma que pode até fazer sentido já que estão ambos inseridos numa hierarquia. Esta ideia acabou por ser afastada, para que o processo se tornasse mais simples para os investigadores.

De seguida surgiram questões relativas à identificação de um ficheiro ou pasta. Por outras palavras, a dúvida residia em saber se seria aquele o ficheiro ou a pasta a ser anotado, uma vez que dentro de diretórios diferentes poderão existir os mesmos nomes de pastas ou de ficheiros. Através da utilização do caminho <sup>8</sup>, por exemplo “InvestigadorA/ProjetoA/Pasta1/Ficheiro1” não existe forma de confundir as anotações.

Por fim foi necessário pensar como iria ser estabelecida a comunicação entre ambas as aplicações. Ficou decidido que a UPBox ficaria responsável por fornecer os seguintes serviços (Mais detalhes sobre os serviços disponibilizados pela UPBox podem ser consultados no Anexo B <sup>9</sup>):

- [http://dendro-dev.fe.up.pt:3000/get\\_userprojects.json?datanoteskey=\[TOKEN\]&uid=\[USER\]](http://dendro-dev.fe.up.pt:3000/get_userprojects.json?datanoteskey=[TOKEN]&uid=[USER]) — Um método GET que retorna todos os projetos aos quais um utilizador tem acesso;
- [http://dendro-dev.fe.up.pt:3000/home/\[USER\]/\[PROJECT\].json?datanoteskey=\[TOKEN\]](http://dendro-dev.fe.up.pt:3000/home/[USER]/[PROJECT].json?datanoteskey=[TOKEN]) — Um método GET que retorna toda a estrutura de diretórios e ficheiros do projeto selecionado por um utilizador;
- [http://dendro-dev.fe.up.pt/datanotes/Special:ExportRDF/\[PATH\]](http://dendro-dev.fe.up.pt/datanotes/Special:ExportRDF/[PATH]) — Um método POST com um pedido de envio de um ficheiro no formato RDF com a anotação de uma pasta ou ficheiro (um exemplo de um ficheiro encontra-se no Anexo B.4 na página 86).

Uma vez criada a extensão DataNotes, seria necessário incluir um componente capaz de modificar o menu lateral esquerdo da base da MW para conseguir colocar a estrutura de diretórios e ficheiros vindo da UPBox. Consultada a lista de *hooks* disponibilizados pela MW e SMW descobriu-se um *hook* capaz de corresponder ao pretendido:

- `$wgHooks['SkinBuildSidebar'][] = 'CategoryTagCloud::renderTagCloudBox';`

Apesar de não ser necessária a criação de ficheiros, como quando se criam páginas especiais, esta metodologia foi abordada de forma a contribuir para a organização do código como ilustra a Figura 6.7. Embora esteja voltada para a integração com a UPBox, este conjunto de ficheiros faz também parte da extensão DataNotes.

Para a implementação do método “*renderTagCloudBox*” procedeu-se à utilização da linguagem *JavaScript* por estar implementada do lado do cliente e permitir uma interação em tempo real. Para a comunicação com o serviço disponibilizado pela UPBox foi utilizado *Asynchronous JavaScript and XML* (AJAX) <sup>10</sup>.

A Figura 6.8 mostra o resultado da implementação deste componente e a interação que é possível obter junto do mesmo.

<sup>8</sup>Do Inglês: Path

<sup>9</sup>Note que: <http://dendro-dev.fe.up.pt> é o URL da máquina virtual do desenvolvimento.

<sup>10</sup><http://www.w3schools.com/ajax/default.ASP>

## Implementação do DataNotes

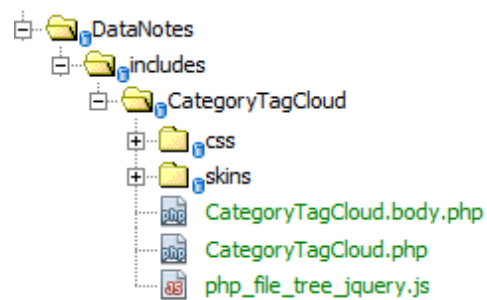


Figura 6.7: Classes da estrutura de diretórios e ficheiros da UPBox no DataNotes

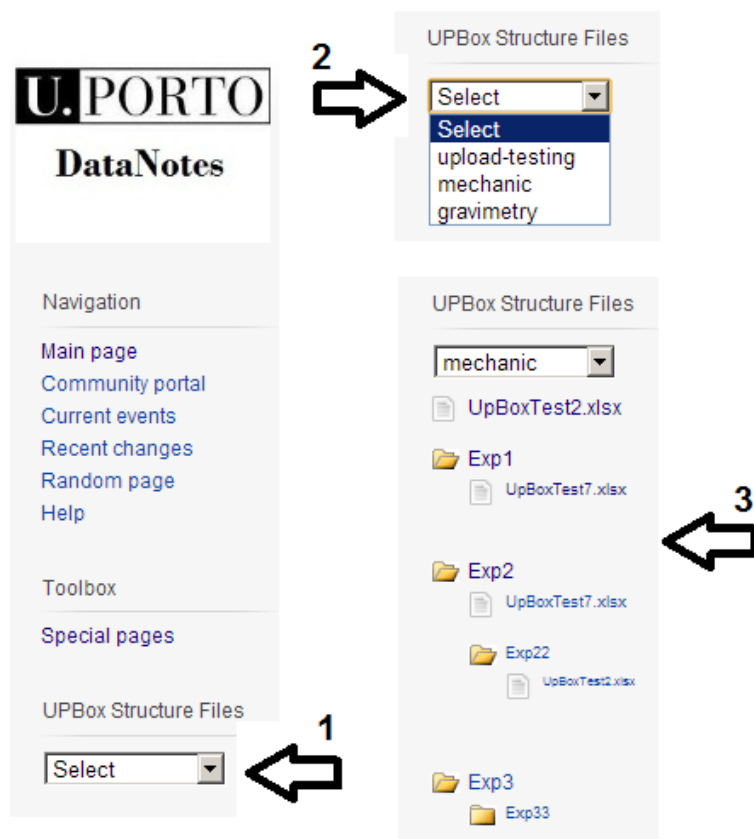


Figura 6.8: Estrutura de diretórios e ficheiros da UPBox no DataNotes

Depois da entrada no DataNotes é automaticamente solicitado ao endereço *http://dendro-dev.fe.up.pt:3000/get\_userprojects.json?datanoteskey=[TOKEN]&uid=[USER]* que disponibilize todos os projetos para os quais o [USER] tem permissões de acesso ao conteúdo. O resultado deste pedido é um ficheiro JSON como mostra a Figura B.1 na página 84. Após a receção deste ficheiro torna-se necessário processá-lo. O atributo “name” é colocado numa *combobox* para permitir que o utilizador selecione o projeto que pretende visualizar. Um pequeno excerto de código encontra-se na Figura B.2 na página 84, onde se mostra implicitamente um pedido AJAX a um URL e o método que processa o ficheiro JSON recebido.

A seleção de um projeto desencadeia um pedido a outro URL *http://dendro-dev.fe.up.pt:3000/home/[USER]/[PROJECT].json?datanoteskey=[TOKEN]*. Este pedido devolve igualmente um ficheiro JSON com todas as pastas, sub-pastas e ficheiros referentes ao projeto selecionado como mostra a Figura B.3 na página 85. Esta informação é processada e colocada numa árvore com auxílio de código em *Javascript*.

A interação nesta árvore permite expandi-la e carregar sobre qualquer pasta, sub-pasta ou ficheiro que se pretenda anotar. Caso não exista ainda anotação sobre o recurso selecionado é criada uma nova página na wiki. Em contrapartida, caso já exista a anotação, esta é apenas desenhada para visualização e se o utilizador assim o desejar poderá posteriormente editá-la.

## 6.5 Resumo e conclusões

Neste capítulo elaborou-se uma abordagem à implementação sobre a plataforma MW e SMW. Foram descritas as noções base que permitiram a criação da extensão DataNotes, bem como os respetivos desafios e detalhes de implementação. Esta implementação decorreu com normalidade e todas as funcionalidades foram implementadas com sucesso.

A integração do DataNotes com a UPBox apresenta uma mais valia a ambos os projetos, no sentido em que ambos integram num processo de curadoria e preservação de dados. Para isso, estabeleceu-se uma comunicação com base em URL que devolvem ficheiros JSON com o que foi pedido.



## Capítulo 7

# Demonstração do DataNotes

Este capítulo de demonstração da extensão DataNotes pretende dar a conhecer o tipo de interfaces que podem ser encontradas para as diversas operações e como podem ser utilizadas.

### 7.1 Introdução

As páginas da SMW são personalizadas de acordo com que o utilizador pretende. A personalização sugerida para a página inicial da wiki DataNotes na Universidade do Porto está ilustrada na Figura 7.1. Pretende-se dar a conhecer ao utilizador as funcionalidades que tem ao seu dispor e que poderá executar.

As permissões de acesso ao conteúdo disponibilizado estão bloqueadas a todos os não utilizadores. Para que possam ter acesso é necessário efetuarem o login na wiki, como mostra a Figura 7.2.

Caso o login seja efetuado como administrador, este terá acesso a todo o conteúdo da wiki. No entanto, se o utilizador pertencer ao grupo dos “*Researchers*” não terá permissões para introdução de esquemas de metadados na wiki.

### 7.2 Gestão dos esquemas de metadados

A introdução de um esquema de metadados pode ser realizada de duas formas: através da criação de propriedades, *template* (conjunto de descritores e respetivas propriedades), formulário (modelo com caixas de texto que contém descritores e propriedades predefinidas) e categoria ou através da criação de uma “classe” que engloba todas as fases de criação anteriores com a diferença de que todas são introduzidos num só momento.

1. **Criação de uma classe** — Ilustrada na Figura 7.3, esta é a funcionalidade mais direta pois permite a criação de um esquema de metadados de uma só vez. Mesmo que as propriedades inseridas já existam na wiki, não se duplicam. Este método é adaptado a todos os curadores que saibam concretamente o esquema de metadados ou a norma que pretende inserir. A criação de uma classe consiste nos seguintes passos:

## Welcome to DataNotes Wiki - University of Oporto

### How can you annotate folders or files?

[\[edit\]](#)

⇒ On the left side of the page you will find the menu **UPBox Structure Files**.

If you want to annotate some folder or file, you will need select the respective project. Then, you can choose the folder or file.

### How can you create vocabularies?

[\[edit\]](#)

Here you can create all vocabulary. It includes fields, properties and the category.

⇒ [Create a class](#)

But if you prefer to do step by step:

⇒ [Create properties](#)

⇒ [Create a template](#)

⇒ [Create a form](#)

⇒ [Create a category](#)

Figura 7.1: Página principal — Wiki DataNotes da Universidade do Porto

## Log in

### Log in

You must have cookies enabled to log in to DataNotes.

Username:

Password:

☐ Remember my login on this browser (for a maximum of 180 days)

[Forgotten your login details?](#)

Figura 7.2: Login — Wiki DataNotes



## Create a class

Enter all the data here to create the properties, template, form and category for a single class. For more options, use the pages [Create a property](#), [Create a template](#), [Create a form](#) and [Create a category](#) instead.

Template name:  ← 1

Output format: ☒ Table ☐ Side infobox ☐ Plain text ☐ Sections

☐ This template can be included multiple times on the page.

Form name:  ← 2

Category name:  ← 3

Field name:	List of values?	Property name:	Type:	Allowed values:
1. <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	Page	<input type="text"/>
2. <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	Page	<input type="text"/>
3. <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	Page	<input type="text"/>
4. <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	Page	<input type="text"/>
5. <input type="text"/>	<input type="checkbox"/>	<input type="text"/>	Page	<input type="text"/>

← 4

← 5

[Add another](#)

[Create](#)

Figura 7.3: Criação de uma classe

(1) Indicação do nome que se pretende dar ao *template*. Normalmente coloca-se aqui o nome do esquema de metadados que se pretende criar;

(2) Indicação do nome que se pretende dar ao formulário. Aqui também se pode colocar o nome do esquema de metadados que se pretende criar;

(3) Indicação do nome que se pretende dar à categoria. Dado que a categoria fica associada a um formulário, também se pode colocar o nome do esquema de metadados que se pretende criar (apenas para consistência de termos);

(4) Indicação do(s) nome(s) que se pretende dar aos descritor(es);

(5) Associação de uma propriedade a cada descritor introduzido, bem como o seu tipo (conjunto de caracteres, data, email, booleano, página, número, entre outros) e poderá também ser indicado uma lista de valores permitidos.

2. **Criação de propriedades, *template*, formulário e categoria** — não tão direta como o item anterior, no entanto traz grandes vantagens para os curadores que tenham a ideia de uma nova propriedade que pretendem vir a utilizar ou até mesmo um subconjunto de descritores que possam vir a ter interesse. É um processo mais lento, mas ao mesmo tempo não é necessário criar obrigatoriamente o esquema de metadados num só momento.

A Figura 7.4 ilustra a página para a criação de uma propriedade. Na página principal da wiki seleciona-se o *link* “*Create property*” que permite o reencaminhamento para esta página. O curador pode introduzir o nome da propriedade que pretende criar no campo “*Property name*” e indicar o respetivo tipo. Os tipos de propriedades disponíveis pela wiki

## Create a property

Property name:  Type:

This property will link to pages that use the form:

If you want this property to only be allowed to have certain values, enter the list of allowed values, separated by commas (if a value contains a comma, replace it with "\","):

Figura 7.4: Criação de uma propriedade

podem ser encontrados na *combobox*. O tipo “página” permite que, sobre o descritor que utilize esta propriedade, possa ser criada uma nova página. Por exemplo, sobre o descritor “nome do país” (Portugal) pode ser adicionada a propriedade do tipo página “tem capital” que permite criar uma página com informação acerca de Lisboa. Os outros tipos resumem-se a conjuntos de caracteres ou conjuntos numéricos. Para finalizar a introdução de uma propriedade seleciona-se o botão “Save page”. Um conjunto de propriedades tem que ser inserido passo a passo.

A Figura 7.5 mostra a página para a criação de um *template*. Na página principal da wiki seleciona-se o link “Create template” que permite o reencaminhamento para esta página. O curador pode introduzir o nome do *template* que pretende criar no campo “Template name” e indicar os respetivos descritores. Para isso, introduz o nome do descritor no campo “Field name” e seleciona a respetiva propriedade semântica que lhe pretende atribuir. Para adicionar mais descritores é necessário carregar no botão “Add Field”. Existe ainda a particularidade de um *template* poder estar associado a qualquer descritor de outro *template* que utilize uma propriedade semântica igual. Por exemplo, o *template* “Livro” tem um descritor “Autor(es)” e respetiva propriedade “foi escrito por” e o *template* “Autores” contém informações pessoais dos autores. É possível criar uma agregação selecionando no campo “Semantic property” a propriedade “foi escrito por” e introduzindo no campo “Title for list”, “Livros produzidos por este autor”. Isto permite que o descritor “Autor(es)” seja um *link* que referencie para esta nova página e mostre todos os livros que foram escritos pelo próprio. Para finalizar a introdução de um *template* seleciona-se o botão “Save page”.

A Figura 7.6 apresenta a página para a criação de um formulário baseado num ou mais *templates*. Na página principal da wiki seleciona-se o link “Create form” que permite o reencaminhamento para esta página. O curador pode introduzir o nome do formulário que pretende criar no campo “Form name” e indicar o *template* que vai usar no campo “Add template”. Um formulário pode integrar vários *templates*. Isto significa que um esquema de metadados pode também integrar vários *templates*. Para finalizar a introdução de um formulário seleciona-se o botão “Save page”.

Por fim, a Figura 7.7 ilustra a página de criação de uma categoria baseada num formulário previamente introduzido. Na página principal da wiki seleciona-se o link “Create

## Demonstração do DataNotes

### Create a template

Template name:

Category defined by template (optional):

Template fields

To have the fields in this template no longer require field names, simply enter the index of each field (e.g. 1, 2, 3, etc.) as the name, instead of an actual name.

Field name:  Display label:  Semantic property:

☐ This field can hold a list of values, separated by commas

Aggregation

To list, on any page using this template, all of the pages that have a certain property pointing to that page, specify the appropriate property below:

Semantic property:

Title for list:

Output format: ☒ Table ☐ Side infobox ☐ Plain text ☐ Sections

Figura 7.5: Criação de um *template*

### Create a form

Form name (the form is usually given the same name as its main template):

Add template:

(You must add at least one template to this form before you can save it.)

Figura 7.6: Criação de um formulário



Figura 7.7: Criação de uma categoria

*category*” que permite o encaminhamento para esta página. O curador pode introduzir o nome da categoria que pretende criar no campo “*Category name*” e indicar o formulário que vai pertencer a esta categoria através do campo “*Default form*”. Para finalizar a introdução de uma categoria seleciona-se o botão “Save page”.

## 7.3 Gestão das anotações

A gestão de anotações pressupõe a anotação, visualização e edição de uma instância de descritores de diversos modelos de metadados.

### 7.3.1 Anotação de um ficheiro ou pasta

Após a introdução de esquemas de metadados na wiki é possível utilizá-los para criar anotações. Através da estrutura de diretórios e ficheiros do menu lateral da wiki, pode navegar-se e seleccionar-se o que se pretende anotar. A Figura 7.8 mostra a página carregada para introdução de uma nova anotação da pasta “Experi222” que é sub-pasta de “Experi22” e “Esperi2” do projeto “*gravimetry*” do utilizador “Ei10124”. Por omissão aparecem os descritores “*Title*”, “*Subject*” e “*Description*” do esquema de metadados *Dublin Core* dado a sua universalidade (poderia também ser colocado o campo “*Author*”). Não é obrigatório o uso destes descritores, ainda que sejam os mais prioritários na descrição de qualquer conjunto de dados.

As Figuras 7.9 e 7.10 apresentam a interação que um utilizador pode experimentar através desta interface. O botão “*Add new vocabulary and descriptors*” permite adicionar uma nova *combobox* com os esquemas de metadados que ainda não estão a ser utilizados para esta anotação. Cada contorno cinzento corresponde a um conjunto de descritores de um esquema de metadados selecionado. Os esquemas de metadados e descritores podem ser inseridos ou apagados da página. O utilizador decide aquilo que melhor descreve os seus dados. Além disso, o utilizador poderá optar por deixar em texto livre o que não conseguir expressar pelos descritores que estão introduzidos na wiki.

### 7.3.2 Visualizar uma anotação

Após a gravação da anotação é gerada uma página de visualização dos conteúdos introduzidos. A Figura 7.11 mostra uma instância de uma anotação com descritores do esquema de metadados *Dublin Core* que, neste caso, corresponde à categoria “DC”. A Figura 7.12 ilustra uma instância

## Demonstração do DataNotes

Create metadata: Ei10124/gravimetry/Experi2/Experi22/Experi222/

ADD NEW VOCABULARY AND DESCRIPTORS

Select ..

Title:

Subject:

Description:

Free Text:

Save page

Figura 7.8: Criação de uma anotação

Create metadata: Ei10124/gravimetry/Experi2/Experi22/Experi222/

ADD NEW VOCABULARY AND DESCRIPTORS

Select ..  
Select ..  
DC  
DD

Free

Save page

Figura 7.9: Introdução de novos esquemas de metadados para a anotação

Create metadata: Ei10124/gravimetry/Experi2/Experi22/Experi222/

ADD NEW VOCABULARY AND DESCRIPTORS

DD[zoomLevels]
✕

GpsAltitude:  ✕

ZoomLevels:  ✕

DC[Data]
✕

Title:  ✕

Language:  ✕

Date:    ✕

Figura 7.10: Introdução de novos descritores para a anotação

de uma anotação com descritores de dois esquemas de metadados com a categoria “DC” e “DD”. Qualquer que seja o número de esquemas de metadados introduzidos, o tipo de visualização será sempre este.

Através da estrutura de diretórios e ficheiros do menu lateral da wiki pode navegar-se e seleccionar-se qualquer pasta ou ficheiro que esteja já anotado. Isto leva a que a página de visualização da instância da anotação abra e se possam visualizar os seus metadados.

### 7.3.3 Edição de uma anotação

Através da estrutura de diretórios e ficheiros do menu lateral da wiki, pode navegar-se e seleccionar-se também qualquer ficheiro que esteja já anotado para permitir a sua edição. Já na página de visualização dos metadados, o botão “*Edit with form*” permite prosseguir para a sua

Ei10124/mechanic/Exp1/

Title	Experience number 1 - Methods
Subject	Mechanic
Description	Automotive service technicians and mechanics, often called service technicians or service techs, inspect, maintain, and repair cars and light trucks.
Format	ISO9067

Main theme: What Automotive Service Technicians and Mechanics Do.

Category: DC

Figura 7.11: Visualização de uma anotação com descritores de um esquema de metadados

Ei10124/mechanic/Exp1/

Title	Experience number 1 - Methods
Subject	Mechanic
Description	Automotive service technicians and mechanics, often called service technicians or service techs, inspect, maintain, and repair cars and light trucks.
Format	ISO9067

zoomLevels 12

Main theme: What Automotive Service Technicians and Mechanics Do.

Categories: DC | DD

Figura 7.12: Visualização de uma anotação com descritores de dois esquemas de metadados

edição. A página de edição da anotação é mostrada e o utilizador pode manipular os esquemas de metadados, descritores e metadados anteriormente escolhidos e preenchidos, como é possível confirmar junto das Figuras 7.13 e 7.14.

## 7.4 Testes

Ao longo do ciclo de desenvolvimento de toda a extensão, vários testes foram realizados. Inicialmente, e após a instalação da MW e SMW, simples tentativas de criação de páginas foram elaboradas de modo a perceber se a instalação tinha sido concluída com sucesso. Numa fase seguinte, ou seja, quando se tentou manipular algum código existente por parte do *Semantic Forms*, verificou-se que devido à forma estrutural que o código se encontrava organizado seria uma difícil solução a adotar. Quando se encontrou uma nova solução capaz de conseguir integrar e reaproveitar parte das ideias existentes no *Semantic Forms*, foram elaborados testes que comprovaram o seu correto funcionamento. No que diz respeito à integração do DataNotes com a UPBox foram testados os pedidos aos endereços URL disponibilizados, bem como o envio do ficheiro RDF após o momento de anotação. Os testes foram elaborados com sucesso. Por fim, após a instalação da extensão DataNotes numa máquina virtual de endereço <http://dendro-dev.fe.up.pt/datanotes> todas as funcionalidades implementadas voltaram a ser testadas. Foram necessários alguns reajustes que posteriormente resultaram no bom funcionamento da extensão.

## 7.5 Resultados e desempenho

Os requisitos da extensão DataNotes visualizados neste capítulo foram desenvolvidos com suporte à plataforma MW e SMW. Como resultado final obteve-se uma plataforma adaptada às necessidades dos investigadores e curadores que pretendem anotar os seus conjuntos de dados. A simplicidade e a interação que o processo de anotação proporciona aos seus utilizadores, permite que não se sintam forçados a descrever campos desconhecidos. A interação com a UPBox através da árvore de estruturas de diretórios e ficheiros faz com que os utilizadores tenham uma visão abrangente de tudo o que está depositado no repositório e que podem anotar.

## Demonstração do DataNotes

Edit metadata: Ei10124/mechanic/Exp1/

ADD NEW VOCABULARY AND DESCRIPTORS

Select ..

Title: Experience number 1 - Methods

Subject: Mechanic

Description: Automotive service technicians and mechanics, often called service technicians or service techs, inspect, maintain, and repair cars and light trucks.

Format: ISO9067

Free Text:  
Main theme: What Automotive Service Technicians and Mechanics Do.

Save page

Figura 7.13: Editar uma anotação

Edit metadata: Ei10124/mechanic/Exp1/

ADD NEW VOCABULARY AND DESCRIPTORS

Select ..

Title: Experience number 1 - Methods

Subject: Mechanic

Description: Automotive service technicians and mechanics, often called service technicians or service techs, inspect, maintain, and repair cars and light trucks.

Format: ISO9067

Free Text:  
Main theme: What Automotive Service Technicians and Mechanics Do.

Save page

Figura 7.14: Adicionar novos descritores a uma anotação em modo de edição



## Capítulo 8

# Conclusões e Desenvolvimentos futuros

Este capítulo final pretende mostrar aquilo que foi desenvolvido com este trabalho, analisar o grau de satisfação dos objetivos, enumerar possíveis trabalhos futuros e realçar as suas contribuições. Todos os projetos são pensados e estruturados antes de serem implementados, mas a verdade é que ao longo do tempo novas ideias vão surgindo acerca do rumo que o trabalho pode levar futuramente. Isto demonstra, acima de tudo, que o trabalho aborda um tema inovador e que ainda desperta vontade de o fazer crescer.

### 8.1 Síntese

Para tornar possível a partilha e reutilização de dados de investigação é necessária a presença de uma estrutura robusta capaz de recolher e preservar os dados de investigação. Além da importância de garantir que os dados permanecem acessíveis a longo prazo, é igualmente importante que sejam também interpretáveis. Quando se tem um conjunto de dados em mão, se não houver nada que o descreva, também não é fácil a sua interpretação por parte de outros investigadores que queiram usá-lo. Os próprios investigadores têm de ser capazes de descrever corretamente os seus dados. De forma a minimizar os recursos necessários à implementação de uma solução de curadoria eficaz, este trabalho focou-se no desenvolvimento de um sistema de anotação colaborativa destinado ao uso por parte dos seus investigadores.

Inicialmente foram elaboradas diversas pesquisas que tinham como objetivo a investigação das potencialidades das tecnologias apropriadas à tarefa proposta e foram analisados e observados alguns projetos desenvolvidos no âmbito dos repositórios de dados científicos. Dos projetos analisados, poucos se centravam na vertente de descrição dos dados de investigação sem que fosse necessária a existência de um curador. Após esta visão geral, a pesquisa focou-se sobre tecnologias wiki. Uma vez que se pretendia um sistema colaborativo, isto é, um sistema onde diversos investigadores pudessem colaborar na anotação de um conjunto de dados, este tipo de tecnologia seria capaz de fornecer algumas funcionalidades base de suporte: login na wiki, gestão de utilizadores, gestão de páginas, entre outros. Uma das soluções encontrada e, posteriormente utilizada

como tecnologia principal foi a *Semantic MediaWiki*. Existem não só muitas referências à utilização desta wiki, como também inúmeras contribuições de extensões que alargam a sua gama de aplicações. Através da análise de algumas dessas extensões foi possível apurar que o *Semantic Forms*, extensão que permite a criação de formulários na wiki, seria um bom ponto de partida para a interação com os investigadores.

Tendo por base a instalação da SMW optou-se pela implementação de uma nova extensão, DataNotes, em linguagem PHP suportada por uma base de dados MySQL. Esta nova extensão seguiu algumas das ideias já desenvolvidas para a extensão *Semantic Forms*, nomeadamente ao disponibilizar a anotação em modo de formulário para o curador e para o investigador. Numa tentativa de reaproveitar aquilo que já estava implementado por parte do *Semantic Forms*, percebeu-se que o código era pouco flexível devido ao elevado número de linhas num só método e respetivas condições lógicas.

As tarefas realizadas consistiram na criação de uma estrutura de diretórios e ficheiros referente a um determinado investigador, para que o investigador possa navegar neles e obter as suas descrições. Foi também implementada a página wiki para a visualização, criação e edição de uma anotação de forma interativa recorrendo à linguagem *Javascript*. A introdução de esquemas de metadados nesta wiki segue uma formalização baseada em propriedades, *templates*, formulários e categorias já previamente implementados na SMW.

Num âmbito geral, os objetivos inicialmente propostos para este projeto foram cumpridos. Para além de outros fatores, a definição de um plano de trabalho estabelecido numa fase inicial contribuiu para o sucesso do alcance dos objetivos. Apesar de terem sido realizadas, pontualmente, alterações na duração de algumas tarefas devido a imprevistos que foram surgindo, o plano estabelecido foi sendo respeitado ao longo do desenvolvimento de todo o projeto.

Em suma, o desenvolvimento desta ideia inovadora foi cumprido. A integração com o projeto UPBox conseguiu que ambos, num ponto de vista de processo de curadoria, respondessem às necessidades de armazenamento, preservação e descrição dos dados.

## 8.2 Perspetivas e sugestões futuras

A ideia da implementação deste trabalho sob uma plataforma wiki levou a que se gerassem novas ideias de desenvolvimentos futuros. A possibilidade de extensão desta plataforma por outros pontos do mundo, a partilha das extensões desenvolvidas por todos aqueles que pretendem também colaborar e a possibilidade deste trabalho ficar conhecido por outras entidades são fortes motivos que levaram à escolha desta plataforma.

A verdade é que são os investigadores os autores dos dados científicos que se pretendem armazenar e anotar. Os investigadores sentem-se desencorajados por terem que se dirigir a alguém durante largas horas para dar conhecimento do que se pode retirar como informação útil daqueles dados. Com isto, acabam por manter os dados nos seus computadores e não os partilham.

Espera-se que esta plataforma venha encorajar os investigadores a serem eles capazes de conduzir este processo; no entanto, para que toda a interação fique concluída, ainda faltam alguns

desafios e interligações, nomeadamente no que diz respeito à fase de aprovação para curadoria. A ligação deste projeto e do projeto UPBox ao DSpace irá permitir fechar o processo de curadoria, ou seja, assim que o curador conclua que os metadados introduzidos para a anotação de uma pasta ou ficheiro estão capazes de assegurar uma boa descrição dos dados, não permite que sejam efetuadas mais alterações sobre os metadados. Um registo da instância de descritores de vários esquemas de metadados fica então guardado.

Seria igualmente interessante que o investigador pudesse fazer parte da partilha e introdução de novos esquemas de metadados na wiki, ainda que fosse exigida uma avaliação por parte de um curador antes de ficar disponível publicamente. Com isto pretende-se que o investigador, grande parte das vezes o que melhor domina a sua área, possa partilhar normas e esquemas de metadados que conheça com um curador. Outra ideia interessante nesta linha de pensamento é o desenvolvimento de uma nova extensão sobre a SMW que permita que os esquemas de metadados sejam importados para o DataNotes e posteriormente fazer com que parte da anotação seja preenchida de forma automática, através da exploração dos conteúdos dos conjuntos de dados a ser anotados.

Admitindo que a adesão a esta extensão seria grande, a implementação de um sistema de gestão de utilizadores com autenticação LDAP seria uma outra sugestão, dado que o projeto UPBox já a integra. Deste modo, a sintonia entre ambas as plataformas funcionaria com níveis de segurança mais desejados.

No desenvolvimento deste projeto procurou-se criar um conceito inovador para a gestão de dados científicos. O processo de anotação de dados que até agora recorria a um curador, com base no DataNotes passou a fazer parte do próprio criador dos dados, o investigador. Dispondo de melhores ferramentas, esperamos que os investigadores possam aperceber-se cedo do valor dos seus dados e que estejam motivados para os descrever; com modelos de metadados mais abertos, é também mais fácil os investigadores encontrarem descritores apropriados para os seus dados.

## Conclusões e Desenvolvimentos futuros

# Referências

- [Bar08] Daniel J. Barrett. *MediaWiki*. O'Reilly Media, Estados Unidos da América, 1st edition, 2008.
- [Bar13] José Barbosa. UPBox — Armazenamento na Nuvem para Dados de Investigação da U. Porto. Master's thesis, 2013.
- [Boa12] DCMI Usage Board. DCMI Metadata Terms. Disponível em <http://dublincore.org/documents/dcmi-terms/>, acedido em 27 de Junho de 2012, 2012.
- [Bor11] Christine L. Borgman. The Conundrum of Sharing Research Data. *SSRN Electronic Journal*, pages 1–40, 2011. doi:10.2139/ssrn.1869155.
- [CB10] Maria Elisabete Catarino e Ana Alice Baptista. Ontologia STAP: Um Vocabulário de Termos de Metadados. Disponível em <http://repositorium.sdum.uminho.pt/handle/1822/11120>, acedido em 16 de Janeiro de 2013, 2010.
- [CER11] CERN. Supporting Open Access Publishing. Disponível em <http://library.web.cern.ch/library/OpenAccess/OpenAccessPolicy.html>, acedido em 23 de Junho de 2012, 2011.
- [Com13] Fedora Commons Inc. Fedora Commons. Disponível em <http://www.fedora-commons.org/>, acedido em 6 de Janeiro de 2013, 2013.
- [Cor12] USA Cornell University Library. DataStAR. Disponível em <http://datastar.mannlib.cornell.edu/>, acedido em 12 de Dezembro de 2012, 2012.
- [DAM13] DAMA International. Data Management International. Disponível em [www.dama.org](http://www.dama.org), acedido em 16 de Janeiro de 2013, 2013.
- [DCC12] DCC. DCC — Digital Curation. Disponível em <http://www.dcc.ac.uk/digital-curation/what-digital-curation>, acedido em 22 de Junho de 2012, 2012.
- [DU11] DISC-UK. DISC-UK Home. Disponível em <http://www.disc-uk.org/>, acedido em 22 de Junho de 2012, 2011.
- [Dur12a] Duraspace. Disponível em <http://www.duraspace.org/>, acedido em 6 de Janeiro de 2013, 2012.
- [Dur12b] DuraSpace. DSpace. Disponível em <http://www.dspace.org/introducing>, acedido em 16 de Junho de 2012, 2012.
- [Dur13] Duracloud. Disponível em <http://duracloud.org/>, acedido em 6 de Janeiro de 2013, 2013.

## REFERÊNCIAS

- [HSS11] Bernhard Haslhofer, Rainer Simon e Robert Sanderson. The Open Annotation Collaboration (OAC) Model. *Arxiv preprint arXiv*., 2011. arXiv:arXiv:1106.5178v1.
- [Hun07] Jane Hunter. (DA3) Metadata Schema Registry. Technical Report May, Australian Commonwealth Department of Education, Science and Training, 2007.
- [JIS09] JISC; UKOLN. UKOLN — Metadata Schema Registry. Disponível em <http://www.ukoln.ac.uk/projects/iemsr/>, acessado em 22 de Junho de 2012, 2009.
- [JoB12] JISC e University of Bath. UKOLN — Metadata. Disponível em <http://www.ukoln.ac.uk/metadata/>, acessado em 22 de Junho de 2012, 2012.
- [KCCr<sup>+</sup>10] Huda Khan, Brian Caruso, Jon Corson-rikert, Dianne Dietrich, Brian Lowe e Albert R Mann Library. Using the Semantic Web approach for Data Curation. Number December, pages 1–12. 6th International Digital Curation Conference DataStaR, 2010.
- [KKPS02] J. Kahan, M.-R. Koivunen, E. Prud’Hommeaux e R.R. Swick. Annotea: an open RDF infrastructure for shared Web annotations. In ELSEVIER Science B.V., editor, *Computer Networks* 39, pages 589–608. 2002.
- [KMH08] Chrysovalanto Kousetti, David E. Millard e Yvonne Howard. A study of ontology convergence in a semantic Wiki. *Proceedings of the 4th International Symposium on Wikis — WikiSym*, page 1, 2008. doi:10.1145/1822258.1822281.
- [Kor12] Yaron Koren. *Working with MediaWiki*. WikiWorks Press, 1st edition, 2012.
- [KVV<sup>+</sup>07] Markus Krötzsch, Denny Vrandečić, Max Völkel, Heiko Haller e Rudi Studer. Semantic Wikipedia. *Journal of Web Semantics* 5, pages 251–261, 2007.
- [MJLP10] Thomas Meilender, Nicolas Jay, Jean Lieber e Fabien Palomares. Semantic wiki engines: a state of the art. *Semantic Web Journal*, 1:1–5, 2010.
- [OMS<sup>+</sup>06] Eyal Oren, Knud Hinnerk M, Simon Scerri, Siegfried Handschuh e Michael Sintek. What are Semantic Annotations? 2006. doi:10.1.1.97.7985.
- [oS07a] University of Southampton. EPrints — Demo. Disponível em [http://www.eprints.org/software/v3/EPrintsv3Presentation\\_small.pdf](http://www.eprints.org/software/v3/EPrintsv3Presentation_small.pdf), acessado em 5 de Julho de 2012, 2007.
- [oS07b] University of Southampton. EPrints — Video. Disponível em <http://www.eprints.org/software/videos.php>, acessado em 5 de Julho de 2012, 2007.
- [PNN<sup>+</sup>07] Andy Powell, Mikael Nilsson, Ambjörn Naeve, Pete Johnston e Thomas Baker. DCMI Abstract Model. Disponível em <http://dublincore.org/documents/abstract-model/>, acessado em 27 de Junho de 2012, 2007.
- [RRFS10] Cristina Ribeiro, Elloy Rodrigues, Eugénia Fernandes e Rui Saraiva. Os repositórios de dados científicos: Estado da Arte. RCAAP. Technical report, Universidade do Porto e Universidade do Minho, Portugal, 2010.
- [RRL11] João Rocha, Cristina Ribeiro e João Correia Lopes. A Data Curation Experiment at U.Porto using DSpace. In *Proceedings of 8th International Conference on Preservation of Digital Objects, iPRES 2011*, pages 224–227, 2011.

## REFERÊNCIAS

- [SBBK08] Sebastian Schaffert, Fran Bry, Joachim Baumeister e Malte Kiesel. Semantic Wikis. *IEEE Software*, 25(4):8–11, July 2008. doi:10.1109/MS.2008.95.
- [Sim12] Simmons. Open Access Directory. Disponível em [http://oad.simmons.edu/oadwiki/Data\\_repositories](http://oad.simmons.edu/oadwiki/Data_repositories), acessado em 11 de Junho de 2012, 2012.
- [Siq12] Jéssica Câmara Siqueira. A noção de folksonomia: uma abordagem terminológica. *TradTerm*, 20:129–145, 2012.
- [SL10] Gail Steinhart e Albert R Mann Library. DataStaR: a data staging repository to support the sharing and publication of research data. In *IATUL Conference*, International Association of Scientific and Technological University Libraries, 2010.
- [SWP03] Thornton Staples, Ross Wayland e Sandra Payette. The Fedora Project — An Open-source Digital Object Repository Management System. *D-Lib Magazine Volume 9 Number 4*, 2003.
- [Uni12a] Universidade do Porto. UPData. Disponível em <http://sciencedata.up.pt/dspace/>, acessado em 20 de Maio de 2012, 2012.
- [Uni12b] University of Essex. How we curate data — The process. Disponível em <http://www.data-archive.ac.uk/curate/process>, acessado em 2 de Julho de 2012, 2012.
- [Uni12c] UK University of Southampton. EPrints. Disponível em <http://www.eprints.org/openaccess/>, acessado em 8 de Julho de 2012, 2012.
- [VKV<sup>+</sup>06] Max Völkel, Markus Krötzsch, Denny Vrandečić, Heiko Haller e Rudi Studer. Semantic Wikipedia. 2006.
- [W3C04a] W3C. OWL Web Ontology Language — Overview. Disponível em <http://www.w3.org/TR/owl-features/>, acessado em 4 de Janeiro de 2013, 2004.
- [W3C04b] W3C. Resource Description Framework (RDF): Concepts and Abstract Syntax. Disponível em <http://www.w3.org/TR/rdf-concepts/>, acessado em 4 de Janeiro de 2013, 2004.
- [W3C08] W3C. Extensible markup language (xml) 1.0 (fifth edition). Disponível em <http://www.w3.org/TR/REC-xml/#sec-origin-goals>, acessado em 4 de Janeiro de 2013, 2008.

## REFERÊNCIAS



## Anexo A

# Requisitos da extensão DataNotes

Tabela A.1: UC1 — Login na wiki DataNotes

<b>ID</b>	UC1
<b>Nome</b>	Login
<b>Descrição</b>	Permite ao utilizador entrar na wiki e ter permissões para gerir as suas anotações, entre outras tarefas.
<b>Ator(es)</b>	Curador, Investigador
<b>Pré condições</b>	Existência do utilizador na base de dados de utilizadores da wiki.
<b>Pós condições</b>	O utilizador entra no sistema.
<b>Fluxo</b>	<ol style="list-style-type: none"><li>1. O utilizador acede ao <i>browser</i> e introduz o respetivo URL para aceder à wiki.</li><li>2. O utilizador introduz o seu nome de utilizador e <i>password</i>.</li><li>3. O utilizador seleciona o botão “Autenticação”.</li><li>3.1. Se o utilizador introduzir o nome de utilizador ou a <i>password</i> incorretos, a wiki lança um alerta “Erro de autenticação”.</li></ol>

Tabela A.2: UC2 — Selecionar ficheiro/pasta na UPBox

<b>ID</b>	UC2
<b>Nome</b>	Selecionar ficheiro/pasta na UPBox
<b>Descrição</b>	Permite ao utilizador selecionar uma pasta ou um ficheiro que pretende anotar.
<b>Ator(es)</b>	Curador, Investigador
<b>Pré condições</b>	O utilizador está autenticado como curador ou investigador. Existência de projetos e respetivo conteúdo na estrutura de diretórios da UP-Box. Existência do ficheiro/pasta que pretende anotar na estrutura de diretórios UP-Box.
<b>Pós condições</b>	—
<b>Fluxo</b>	1. O utilizador seleciona o projeto que pretende visualizar na sua estrutura de diretórios. 2. A árvore referente a esse projeto é apresentada do lado esquerdo no menu lateral da wiki. 3. O utilizador navega entre todos os ficheiros/pastas e seleciona aquele que pretende visualizar/anotar/editar/eliminar. 3.1. Se não existir nenhuma anotação para aquele ficheiro/pasta será criada uma nova página para anotação. 3.2. Se já existir uma anotação para aquele ficheiro/pasta será possível visualizar o conteúdo da mesma e posteriormente editar ou eliminar.

Tabela A.3: UC3 — Inserir uma anotação

<b>ID</b>	UC3
<b>Nome</b>	Inserir uma anotação
<b>Descrição</b>	Permite ao utilizador anotar um conjunto de dados ou mesmo uma estrutura de diretórios que contém conjuntos de dados.
<b>Ator(es)</b>	Curador, Investigador
<b>Pré condições</b>	O utilizador está autenticado como curador ou investigador. Existência de esquemas de metadados na wiki.
<b>Pós condições</b>	Um página nova é criada na wiki, com a informação da anotação realizada.
<b>Fluxo</b>	1. O utilizador escolhe um ou vários esquemas de metadados a utilizar na anotação. 2. Uma lista de descritores do(s) esquema(s) de metadados que vão sendo selecionados são mostrados. 3. O utilizador preenche os descritores escolhidos. 3.1. Se o utilizador pretender remover algum esquema e respetivos descritores anteriormente selecionado, pode fazê-lo no botão respetivo para apagar(X). 3.2. Se o utilizador pretender remover algum descritor anteriormente selecionado, pode fazê-lo no botão respetivo para apagar(X). 4. O utilizador seleciona o botão “Gravar” assim que a sua anotação tiver sido concluída.

Tabela A.4: UC4 — Visualizar uma anotação

<b>ID</b>	UC4
<b>Nome</b>	Visualizar uma anotação
<b>Descrição</b>	Permite ao utilizador visualizar ficheiros/pastas que foram previamente anotados.
<b>Ator(es)</b>	Curador, Investigador
<b>Pré condições</b>	O utilizador está autenticado como curador ou investigador.
	Existência da anotação que se pretende visualizar na wiki.
<b>Pós condições</b>	—
<b>Fluxo</b>	1. O utilizador visualiza uma página com o conteúdo da anotação previamente elaborada.
	1.1. Se o utilizador tiver interessado em visualizar o conteúdo semântico associado, poderá explorar os links que lhe são disponibilizados nos metadados.

Tabela A.5: UC5 — Editar uma anotação

<b>ID</b>	UC5
<b>Nome</b>	Editar uma anotação
<b>Descrição</b>	Permite ao utilizador editar uma anotação previamente criada.
<b>Ator(es)</b>	Curador, Investigador
<b>Pré condições</b>	O utilizador está autenticado como curador ou investigador.
	Existência de esquemas de metadados na wiki.
	Existência do ficheiro/pasta selecionado para editar.
<b>Pós condições</b>	A página da anotação é editada.
<b>Fluxo</b>	1. Sobre a visualização do ficheiro/pasta para o qual pretende editar a anotação, o utilizador seleciona “Editar com formulário”.
	2. O utilizador escolhe um ou vários esquemas de metadados ou descritores que pretende acrescentar ou retirar na anotação.
	2.1. Se o utilizador escolhe novos esquemas de metadados, então uma lista de descritores do(s) esquema(s) que vão sendo selecionados é mostrada.
	2.2. Se o utilizador escolhe novos descritores, novos campos vão sendo mostrados para anotar.
	2.3. Se o utilizador escolhe retirar um esquemas de metadados, então esse esquemas desaparece da anotação (no entanto pode voltar a colocá-lo).
	2.4. Se o utilizador escolhe retirar um descritor, então esse descritor desaparece da anotação (no entanto pode voltar a colocá-lo).
	3. Após a edição daquilo que pretende, o utilizador seleciona o botão “Gravar”.

## Requisitos da extensão DataNotes

Tabela A.6: UC6 — Eliminar uma anotação

<b>ID</b>	UC6
<b>Nome</b>	Eliminar uma anotação
<b>Descrição</b>	Permite ao utilizador eliminar uma anotação previamente criada.
<b>Ator(es)</b>	Curador, Investigador
<b>Pré condições</b>	O utilizador está autenticado como curador ou investigador. Existência de vocabulários na wiki.
<b>Pós condições</b>	Um página nova é criada na wiki, com a informação da anotação realizada.
<b>Fluxo</b>	1. Sobre a visualização do ficheiro/pasta da anotação que pretende editar, o utilizador seleciona “Eliminar”. 2. O utilizador tem de confirmar se pretende na verdade eliminar. 2.1. O utilizador pode justificar o motivo da remoção.

Tabela A.7: UC7 — Gerir grupos

<b>ID</b>	UC7
<b>Nome</b>	Gerir grupos
<b>Descrição</b>	Permite ao utilizador efetuar a gestão de permissões de novos utilizadores da wiki.
<b>Ator(es)</b>	Curador
<b>Pré condições</b>	O utilizador está autenticado como curador. Existência do utilizador na wiki.
<b>Pós condições</b>	O utilizador fica atribuído a (um) novo(s) grupo(s).
<b>Fluxo</b>	1. O utilizador acede através do campo pesquisa da wiki: “Special:UserRights” 2. O utilizador introduz o nome do utilizador que pretende gerir. 2.1. Se o nome do utilizador inserido não existir, a wiki envia um alerta: “Não existe um utilizador com o nome [nomeUtilizador]. Verifique o nome que introduziu.” 3. O utilizador pode seleccionar o(s) grupo(s) que pretende que o utilizador pertença. 4. Por fim, o utilizador seleciona o botão “Gravar grupos do utilizador”

Tabela A.8: UC8 — Inserir um esquema de metadados

<b>ID</b>	UC8
<b>Nome</b>	Inserir um esquema de metadados
<b>Descrição</b>	Processo de inserção de um esquema de metadados na wiki.
<b>Ator(es)</b>	Curador
<b>Pré condições</b>	O utilizador está autenticado como curador.
<b>Pós condições</b>	Um novo esquema de metadados é inserido na wiki.
<b>Fluxo</b>	1. O utilizador insere propriedade(s) na wiki. (UC9) 2. O utilizador insere um <i>template</i> na wiki. (UC10) 3. O utilizador cria um formulário na wiki. (UC11) 4. O utilizador cria uma categoria na wiki. (UC12)

Tabela A.9: UC9 — Inserir propriedade

<b>ID</b>	UC9
<b>Nome</b>	Inserir propriedade
<b>Descrição</b>	Inserção de uma propriedade na wiki.
<b>Ator(es)</b>	Curador
<b>Pré condições</b>	O utilizador está autenticado como curador. A propriedade a inserir não faz parte da wiki.
<b>Pós condições</b>	Uma propriedade é introduzida na wiki.
<b>Fluxo</b>	1. Na página principal da wiki, o utilizador encontra um link: “ <i>Create Property</i> ” sobre o qual deve seleccionar. 2. O utilizador introduz o nome da propriedade e indica o seu tipo. Os outros campos poderão ser ignorados no contexto de criação de um esquema de metadados. 3. O utilizador selecciona o botão “Gravar Página” para gravar a propriedade.

Tabela A.10: UC10 — Inserir *template*

<b>ID</b>	UC10
<b>Nome</b>	Inserir <i>template</i>
<b>Descrição</b>	Inserção de um <i>template</i> na wiki.
<b>Ator(es)</b>	Curador
<b>Pré condições</b>	O utilizador está autenticado como curador.
<b>Pós condições</b>	Um <i>template</i> é introduzida na wiki. Se o nome do <i>template</i> introduzido já existir na wiki, é substituído.
<b>Fluxo</b>	1. Na página principal da wiki, o utilizador encontra um link: “ <i>Create Template</i> ” sobre o qual deve seleccionar. 2. O utilizador introduz o nome do <i>template</i> e todos os descritores que pretende introduzir com as respectivas propriedades. Os outros campos poderão ser ignorados no contexto de criação de um esquema de metadados. 2.1 Se o utilizador quiser inserir mais do que um descritor, deverá seleccionar no botão “Adicionar campo”. 2.2 Se o utilizador pretender remover algum campo introduzido por engano, deverá seleccionar o botão “Apagar”. 3. O utilizador selecciona o botão “Gravar Página” para gravar o <i>template</i> .

Tabela A.11: UC11 — Criar formulário

<b>ID</b>	UC11
<b>Nome</b>	Criar formulário
<b>Descrição</b>	Criação de um formulário na wiki.
<b>Ator(es)</b>	Curador
<b>Pré condições</b>	O utilizador está autenticado como curador.
	O formulário a inserir não faz parte da wiki.
<b>Pós condições</b>	Um formulário é introduzido na wiki.
<b>Fluxo</b>	1. Na página principal da wiki, o utilizador encontra um link: “ <i>Create Form</i> ” sobre o qual deve seleccionar.
	2. O utilizador introduz o nome do formulário e indica a que <i>template</i> faz referência.
	3. O utilizador selecciona o botão “Gravar Página” para gravar o formulário.

Tabela A.12: UC12 — Criar categoria

<b>ID</b>	UC12
<b>Nome</b>	Criar categoria
<b>Descrição</b>	Criação de uma categoria na wiki.
<b>Ator(es)</b>	Curador
<b>Pré condições</b>	O utilizador está autenticado como curador.
	A categoria a inserir não faz parte da wiki.
<b>Pós condições</b>	Uma categoria é introduzida na wiki.
<b>Fluxo</b>	1. Na página principal da wiki, o utilizador encontra um link: “ <i>Create Category</i> ” sobre o qual deve seleccionar.
	2. O utilizador introduz o nome da categoria e indica o formulário que esta categoria pretende usar. Os outros campos poderão ser ignorados no contexto de criação de um vocabulário.
	3. O utilizador selecciona o botão “Gravar Página” para gravar a categoria.

## Anexo B

# Interação UPBox e DataNotes

### B.1 API UPBox

As seguintes tabelas contêm a especificação da API que a UPBox disponibiliza ao DataNotes para estabelecer a sua comunicação:

Tabela B.1: API UPBox — Receber ficheiro de anotação do DataNotes

<b>Nome</b>	/home
<b>Descrição</b>	Obter ficheiros e pastas para o caminho especificado.
<b>Estrutura URL</b>	<i>http://dendro-dev.fe.up.pt/home/&lt;USER&gt;/&lt;PROJECT&gt;/(&lt;FOLDER&gt;)</i>
<b>Método</b>	GET
<b>Parâmetros</b>	<i>datanoteskey</i> : chave partilhada com o DataNotes
<b>Retorno</b>	A estrutura de diretórios e ficheiros em XML ou JSON.
<b>Erros</b>	Bad request: 400, "Bad Request".
	Bad auth_token: 401, "Unauthorized".
	No directory: 404, "Not Found".

Tabela B.2: API UPBox — Obter projetos de um utilizador

<b>Nome</b>	/get_userprojects
<b>Descrição</b>	Obter todos os projetos referentes a um utilizador.
<b>Estrutura URL</b>	<i>http://dendro-dev.fe.up.pt/get_userprojects</i>
<b>Método</b>	GET
<b>Parâmetros</b>	<i>uid</i> : identificação do utilizador
	<i>datanoteskey</i> : chave partilhada com o DataNotes
<b>Retorno</b>	Um array com os projetos de um utilizador em XML ou JSON.
<b>Erros</b>	Bad auth_token: 401, "Unauthorized".
	Bad auth_token: 404, "Not Found".: O utilizador não foi encontrado.

Tabela B.3: API UPBox — Envio de um ficheiro com as anotações em RDF

<b>Nome</b>	/Special:ExportRDF
<b>Descrição</b>	Pedido de um ficheiro RDF com as anotações de uma pasta ou ficheiro.
<b>Estrutura URL</b>	<i>http://dendro-dev.fe.up.pt/datanotes/index.php/Special:ExportRDF/&lt;PATH&gt;</i>
<b>Método</b>	PUT
<b>Parâmetros</b>	-
<b>Retorno</b>	-
<b>Erros</b>	Bad auth_token: 404, "Not Found".: Não encontrado.

## B.2 Projetos e estrutura de ficheiros de um utilizador

```
[
  {
    "created_at": "2012-12-06T17:37:10Z",
    "desc": "dsfgasg",
    "id": 1,
    "name": "upload-testing",
    "path": "ei08036/upload-testing/",
    "updated_at": "2012-12-06T17:37:10Z"
  },
  {
    "created_at": "2012-12-09T16:01:48Z",
    "desc": "",
    "id": 2,
    "name": "mechanic",
    "path": "ei10124/mechanic/",
    "updated_at": "2012-12-09T16:01:48Z"
  },
  {
    "created_at": "2012-12-18T17:24:07Z",
    "desc": "",
    "id": 3,
    "name": "gravimetry",
    "path": "ei10124/gravimetry/",
    "updated_at": "2012-12-18T17:24:07Z"
  }
]
```

Figura B.1: Resposta ao pedido sobre o URL que devolve os projetos de um utilizador

```
$(document).ready(function() {
  $.ajax({
    type: "GET",
    url: "http://dendro-dev.fe.up.pt:3000/get_userprojects.json?datanoteskey=" + $token + "&uid=" + $user,
    dataType: "json",
    success: function(data) { ... }
  });
});
```

Figura B.2: Código do pedido sobre o URL que devolve os projetos de um utilizador



```
{
  "id":5,
  "path":"ei10124/mechanic/",
  "name":"mechanic",
  "parent_id":4,
  "created_at":"2012-12-09T16:01:48Z",
  "updated_at":"2012-12-09T16:01:48Z",
  "files":[
    {
      "content_type":"application/vnd.openxmlformats-officedocument.spreadsheetml.sheet",
      "created_at":"2012-12-17T18:50:51Z",
      "file_size":8831,
      "folder_id":5,
      "id":40,
      "name":"UpBoxTest2.xlsx",
      "path":"ei10124/mechanic/",
      "uid":"f9a2de40048a273b1105a271379ee367b681fa75",
      "updated_at":"2012-12-17T18:50:51Z"
    }
  ],
  "folders":[
    {
      "id":6,
      "path":"ei10124/mechanic/Exp1/",
      "name":"Exp1",
      "parent_id":5,
      "created_at":"2012-12-09T16:02:10Z",
      "updated_at":"2012-12-09T16:02:10Z",
      "files":[
        {
          "content_type":"application/vnd.openxmlformats-officedocument.spreadsheetml.sheet",
          "created_at":"2012-12-18T10:53:17Z",
          "file_size":8831,
          "folder_id":6,
          "id":43,
          "name":"UpBoxTest7.xlsx",
          "path":"ei10124/mechanic/Exp1/",
          "uid":"e3af8a9a6168f5a9a3be207ed01ceb248b562d01",
          "updated_at":"2012-12-18T10:53:17Z"
        }
      ],
      "folders":[]
    }
  ],
  "folders":[]
},
(...)
```

Figura B.3: Resposta ao pedido sobre o URL que devolve a estrutura de um projeto

## B.3 Ficheiro RDF DataNotes

(...)

```
<swivt:Subject rdf:about=
"http://dendro-dev.fe.up.pt/datanotes/index.php/Special:ExportRDF/datanotes/index.php/Special:
URIResolver/Ei10124/mechanic/Exp1/">
<rdf:type rdf:resource=
"http://dendro-dev.fe.up.pt/datanotes/index.php/Special:ExportRDF/datanotes/index.php/Special:
URIResolver/Category-3ADC"/>
<rdf:type rdf:resource=
"http://dendro-dev.fe.up.pt/datanotes/index.php/Special:ExportRDF/datanotes/index.php/Special:
URIResolver/Category-3ADD"/>
<rdfs:label>Ei10124/mechanic/Exp1/</rdfs:label>
<swivt:page rdf:resource=
"http://dendro-dev.fe.up.pt/datanotes/index.php/Ei10124/mechanic/Exp1/" />
<rdfs:isDefinedBy rdf:resource=
"http://dendro-dev.fe.up.pt/datanotes/index.php/Special:ExportRDF/Ei10124/mechanic/Exp1/" />
<swivt:wikiNamespace rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">0
</swivt:wikiNamespace>
<property:HasDescription rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
Automotive service technicians and mechanics, often called service technicians or service
techs, inspect, maintain, and repair cars and light trucks.
</property:HasDescription>
<property:HasFormat rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ISO9067
</property:HasFormat>
<property:HasSubject rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Mechanic
</property:HasSubject>
<property:HasTitle rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Experience number
1 - Methods</property:HasTitle>
<property:HasZoom rdf:datatype="http://www.w3.org/2001/XMLSchema#double">12</property:HasZoom>
<swivt:wikiPageModificationDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
2013-01-18T01:39:03Z</swivt:wikiPageModificationDate>
<property:Modification_date-23aux rdf:datatype="http://www.w3.org/2001/XMLSchema#double">
2456310.5687847</property:Modification_date-23aux>
<swivt:wikiPageSortKey rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
Ei10124/mechanic/Exp1/</swivt:wikiPageSortKey>
</swivt:Subject>
```

(...)

Figura B.4: Ficheiro RDF do DataNotes — Instância de metadados

## B.4 Diagrama de interação UPBox e DataNotes

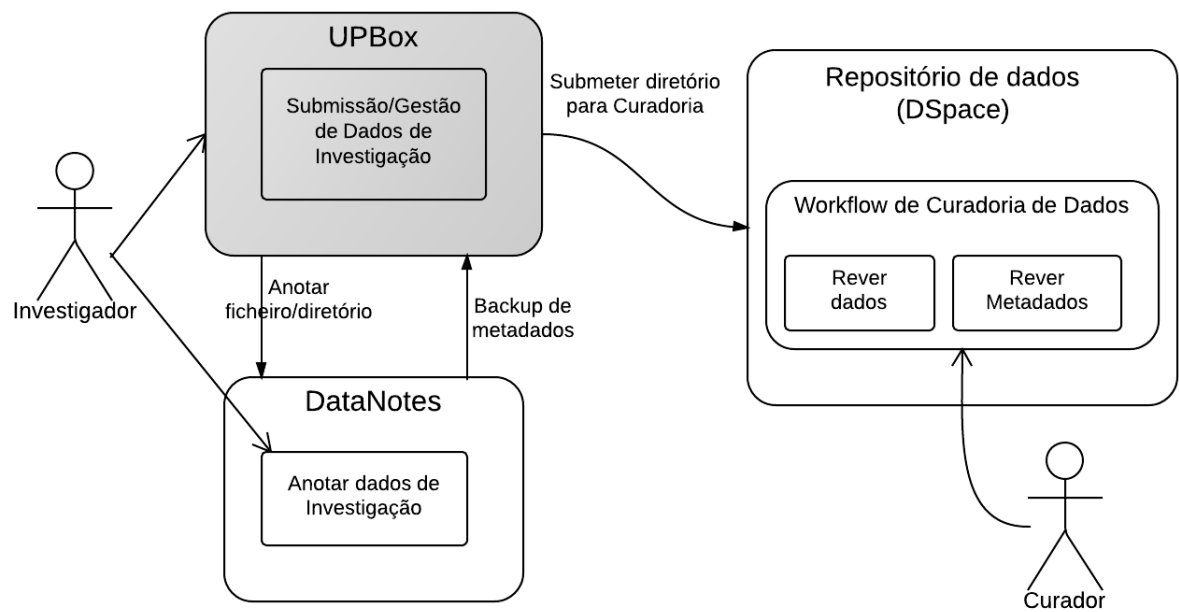


Figura B.5: Diagrama de interação UPBox e DataNotes (in [Bar13])



## Anexo C

# Guia de Instalação

Após a implementação de todo o código, inicialmente testado localmente, foi necessário colocá-lo numa máquina virtual online disponibilizada pela faculdade.

Admite-se que:  $\$DN = /var/www/datanotes$

URL: <http://dendro-dev.fe.up.pt/datanotes>

1. Na pasta WWW da máquina virtual foi necessário criar uma sub pasta denominada por *datanotes*, local onde fica alojado tudo aquilo que é necessário para executar o projeto;  
> `sudo mkdir $DN`
2. Fazer a instalação/atualização de:  
Apache;  
> `sudo apt-get upgrade apache2`  
PHP 5.3.2 ou mais recente;  
> `sudo apt-get upgrade php5`  
MySQL 5.0.2 ou mais recente.  
> `sudo apt-get upgrade mysql-server mysql-client`
3. Fazer o download da MediaWiki 1.19.2;  
> `sudo wget http://download.wikimedia.org/mediawiki/1.19/mediawiki-1.19.2.tar.gz`
4. Descompactar a pasta recebida no download;  
> `sudo tar xvf mediawiki-*.tar.gz`
5. Colocar todos os itens da pasta descompactada no diretório *datanotes*;  
> `sudo mv mediawiki-1.19.2/* $DN`

6. Remover o ficheiro compactado para libertar espaço;  
> `sudo rm mediawiki-1.19.2.tar.gz`
7. Remover a pasta vazia, após a exportação dos seus ficheiros para a raiz;  
> `sudo rm -rf mediawiki-1.19.2`
8. Alterar as permissões da pasta “mw-config”, na pasta raiz *datanotes*;  
> `sudo chmod 777 mw-config/ $DN`
9. Colocar no browser o seguinte endereço:  
> `http://dendro-dev.fe.up.pt/datanotes/install.php`
10. Proceder à instalação da MediaWiki, seguindo todos os passos que são indicados. Estes passos incluem também a criação de uma base de dados, pelo que se deve registar todos os nomes atribuídos e caso se aplique, as credenciais;
11. No final é gerado um ficheiro “LocalSettings.php”, o qual deve ser guardado e colocado na raiz da pasta *datanotes/* (\$DN) com permissões de acesso de administrador. Ou seja, inicialmente coloca-se na pasta raiz do utilizador e posteriormente efetua-se os seguintes comandos;  
> `sudo mv LocalSettings.php $DN`
12. Fazer o *download* da *Semantic MediaWiki* 1.7.1 <sup>1</sup> para o computador;
13. O ficheiro já descompactado é colocado na pasta raiz do utilizador e posteriormente efetua-se os seguintes comandos;  
> `sudo mv SemanticMediaWiki/ $DN/extensions/`  
> `sudo mv Validator/ $DN/extensions/`
14. O ficheiro “LocalSettings.php” necessita de ser alterado. Para isso é necessário editar esse ficheiro;  
> `sudo nano $DN/LocalSettings.php`
15. Acrescentar estas três linhas no fim do ficheiro;  
> `require_once( “$IP/extensions/Validator/Validator.php” );`  
> `include_once( “$IP/extensions/SemanticMediaWiki/SemanticMediaWiki.php” );`  
> `enableSemantics( 'http://dendro-dev.fe.up.pt/datanotes' );`
16. Colocar todos as pastas e ficheiros da raiz *datanotes/* (\$DN) com permissões de administrador (caso contrário podem ocorrer erros de instalação);  
> `sudo chown -R www-data $DN`
17. Proceder à instalação da *Semantic MediaWiki*, seguindo o passo (4) que é indicado nesta página: <http://semantic-mediawiki.org/wiki/Help:Installation>;

---

<sup>1</sup>[http://semantic-mediawiki.org/wiki/Semantic\\_MediaWiki\\_1.7.1](http://semantic-mediawiki.org/wiki/Semantic_MediaWiki_1.7.1)

18. Fazer o *download* do Semantic Forms <sup>2</sup> para o computador;
19. O ficheiro já descompactado é colocado na pasta raiz do utilizador e posteriormente efetuam-se os seguintes comandos;  
> `sudo mv SemanticForms/ $DN/extensions/`
20. O ficheiro “LocalSettings.php” necessita de ser alterado. Para isso é necessário editar esse ficheiro;  
> `sudo nano $DN/LocalSettings.php`
21. Acrescentar esta linha no fim do ficheiro;  
> `include_once( “$IP/extensions/SemanticForms/SemanticForms.php” );`
22. Colocar todos as pastas e ficheiros da raiz datanotes/ para permissões de administrador (caso contrário podem ocorrer erros de instalação);  
> `sudo chown -R www-data $DN`
23. Fazer o *download* do DataNotes para o computador;
24. O ficheiro já descompactado é colocado na pasta raiz do utilizador e posteriormente efetuam-se os seguintes comandos;  
> `sudo mv DataNotes/ $DN/extensions/`
25. O ficheiro “LocalSettings.php” necessita de ser alterado. Para isso é necessário editar esse ficheiro;  
> `sudo nano $DN/LocalSettings.php`
26. Acrescentar esta linha no fim do ficheiro;  
> `include_once( “$IP/extensions/DataNotes/DataNotes.php” );`
27. Colocar todos as pastas e ficheiros da raiz datanotes/ para permissões de administrador (caso contrário podem ocorrer erros de instalação);  
> `sudo chown -R www-data $DN`
28. Testar todo o funcionamento num *browser*;

---

<sup>2</sup>[http://www.mediawiki.org/wiki/Extension:Semantic\\_Forms/Download\\_and\\_installation](http://www.mediawiki.org/wiki/Extension:Semantic_Forms/Download_and_installation)